

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 11:27:40 ; Search time 8363.76 Seconds  
(without alignments)  
1360.163 Million cell updates/sec

Title: US-09-422-999-17  
Perfect score: 3013  
Sequence: 1 gatccagcagagatgtgat.....ccacaatttcaaaatgccc 3013

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est34:\*  
80: gb\_est35:\*  
81: gb\_est36:\*  
82: gb\_est37:\*  
83: em\_est38:\*  
84: em\_est39:\*  
85: em\_est40:\*  
86: em\_est41:\*  
87: gb\_est42:\*  
88: gb\_est43:\*  
89: gb\_est44:\*  
90: gb\_est45:\*  
91: gb\_est46:\*  
92: em\_est47:\*  
93: em\_est48:\*  
94: em\_est49:\*  
95: em\_est50:\*  
96: em\_est51:\*  
97: em\_est52:\*  
98: em\_est53:\*  
99: gb\_est54:\*  
100: gb\_est55:\*  
101: em\_est56:\*  
102: gb\_est57:\*  
103: gb\_est58:\*  
104: gb\_est59:\*  
105: gb\_est60:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	541.4	18.0	675	64	AL119269	AL119269 DKFZ761M
2	527	17.5	600	42	AI099199	AI099199 uc40c07.y

```
3 402.6 13.4 409 64 AL119765 AL119765 DKFZp761D
4 387.2 12.9 446 22 R59283
5 342.4 11.4 541 69 AU079207
6 325.8 10.8 411 64 AM047433
7 315.2 10.5 361 33 AA388005
8 312.8 10.4 317 41 AT003002
9 280.6 9.3 343 64 AL120018
10 268 8.9 268 43 AL1205130
11 263.2 8.7 450 61 AI842094
12 239.8 8.0 277 48 AI608196
13 220.4 7.3 441 45 AI386336
14 215.2 7.1 540 62 AI909224
15 163.6 5.4 44 69 AIW139646
16 158.8 5.3 583 35 AA557278
17 148.4 4.9 413 64 AM084109
18 141.8 4.7 952 80 CNS014CO
19 141.4 4.7 957 28 CNS012W6
20 133.4 4.4 267 78 AA087876
21 127 4.2 538 39 AA890583
22 112.4 3.7 365 23 R88026
23 101.4 3.4 513 43 AI196680
24 100.6 3.3 536 82 AQ717541
25 91.4 3.0 276 22 R47818
26 90.2 3.0 444 61 AI844405
27 89.8 3.0 504 62 AI876724
28 88.8 2.9 378 37 AA687472
29 82.8 2.7 459 31 AA291192
30 82.8 2.7 484 30 AA236800
31 82.8 2.7 493 39 AA890211
32 81.8 2.7 432 33 AA453497
33 81.2 2.6 536 41 AI228074
34 79.4 2.6 593 61 AI830871
35 78.6 2.6 435 41 AI018543
36 73.4 2.4 175 41 AI060273
37 72.8 2.4 459 43 AI227485
38 72.4 2.4 530 42 AI137215
39 72.2 2.4 454 38 AA815190
40 67.4 2.2 408 42 AI137445
41 66.8 2.2 300 30 AA236846
42 66.2 2.2 528 37 AA707689
43 66 2.2 541 62 AI875340
44 65.4 2.2 457 25 W15550
45 64.4 2.1 569 49 AI663371
```

## ALIGNMENTS

```
RESULT 1
LOCUS AL119269 675 bp mRNA EST 27-SEP-1999
DEFINITION DKFZp761M0513_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
ACCESSION AL119269
VERSION DKFZp761M0513.5, mRNA sequence.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Ansoerge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
EST (Ansoerge, Benes, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246636.
Contact: Ansoerge W
MIPS
Am Klopferersplitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL within the cDNA sequencing consortium of the
```

German Genome Project.  
No sl sequence available.  
This clone is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..675

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp761M0513"  
/clone\_1ib="761 (synonym: hamy2)"  
/tissue\_type="amygdala"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"  
BASE COUNT 179 a 172 c 158 g 166 t  
ORIGIN

Query Match 18.0%; Score 541.4; DB 64; Length 675;  
Best Local Similarity 99.1%; Pred. No. 2.6e-143; Indels 4; Gaps 3;  
Matches 576; Conservative 0; Mismatches 1;

```
QY 1 gatcagcgaagatgtgatataatctcactcgaactgaagaagttaagctttgaga 60
|||||
DB 95 GATCCAGCGAAGATGTGATATATCTTCACCTGACCTGAAGAGTTAAGCTTTGAGA 154
61 aattcaaccgaatctcctcatcagattgcttatgtgtattatgagaactcggaaa 120
|||||
DB 155 AATTTTACCCTTCTTCATCATCATGATTTGCTTATGTATATGAAATCTGGAAA 214
121 agggataataattattcgcacagtgatata-tggaacaactcgtatgctg-ccggcag 178
|||||
DB 215 AGGAATATACATATTTCGCGAGGTGATTTGGAACAACTGTATGCTCTGCGCAG 274
179 ggtcttggatgttaagatattcgaacagcagatcaccaggaagtcgtgacatctgta 238
|||||
DB 275 GGTCTTGGATGTTTAAGTATGACACGACGATCACCAGATGCTGTGACCATCTGTA 334
239 cccttggaattggagcgccttggagagtcacattcttgacaacacaccccgcatgcaa 298
|||||
DB 335 CCTTGGGAATTTGGAGCGCTTGGAGAGTTCATTCCTTGGAACAACACCCCGCATGCAA 394
299 ccactgttaacaggaagacagtgactgtcgcgcatcagacgaagaagactccaaggcac 358
|||||
DB 395 CCATCGTTACCAAGGAGACGAGTGAACCTGCTCGCATCAAGAAAGACTTCAAGGCAC 454
359 tatggagaagaatatacagatataatgacagacttgcctcctcctatggtattatgg 418
|||||
DB 455 TATGGAGAAATATCGACAGTATATGCGAGACTTGTGCTCCTCTTATGTTATG 514
419 aaacgggcttaacaatgacagatctcgtgacaaggaagacacac--ctcatgaaactc 476
|||||
DB 515 AAACGGGCTCTTAACAATGACAGATTCTTCGACAAAGGAGAACACCTTCATTGAACCTC 574
477 acgttcctctgtcgtctgtaacacattacacaaagtgccctcaagaagaatccccaag 536
|||||
DB 575 ACCTTCCTCTGCTGCTGCTACACCATTAACCAAGTCTCCTTACAGAGAAATCTCAGAG 634
537 ctggaanaatttaccgaatgcatctctcctcagagcact 577
|||||
DB 635 CTGGAATAATTATGCAATGCAATGCAATCTCTCGACACCACT 675
```

	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 600) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Schellenberg,K., Kuchaba,T., Lacy,M., Le,M., Martin,J., Morris,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The WashU-HMHI Mouse EST Project Unpublished (1996) On Apr 18, 1995 this sequence version replaced gi:775315. Contact: Marra M/Mouse EST Project WashU-HMHI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MGJ:931184 Seq primer: custom primer used High quality sequence stop: 511. Location/Qualifiers 1..600 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1482828" /clone_lib="Sugano mouse liver mlia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: liver; Vector: pME18S-FLJ3; Site_1: DraIII (CACGTGTC); Site_2: DraII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCCTACTGC], digested and cloned into distinct draIII sites of the pME18S-FLJ3 vector (5' site CACGTGTC, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTTAAGACTGCG and 3' end primer CGACCTGAGCTCGAGCA." BASE COUNT     172 a       134 c       153 g       140 t           1 others ORIGIN
Query Match	17.5%; Score 527; DB 42; Length 600;
Best Local Similarity	92.3%; Pred. No. 3.2e-139;
Matches 554; Conservative	0; Mismatches 46; Indels 0; Gaps 0
Dy	844 tgatgagaggtcccgagaacacatgctgtgcgtcaacagataggccccgagcaccacat 903       1 TGATGAAGAAGCTTCAGGAAACCATTGTGCTGCTCTACAGATGGGCCCTTGACGCCACAT 60
Oy	904 gagagatgcccttcggaaacacacctgcccagagagactgtgatgatcctaagattacta 963       61 GAGAAATGATCTCCGGAACAACACCTTGCCAGAGACTGTGGATGAGACTTASAGATTATCTA 120
Oy	964 tgaagagcttctcatatattaagaaccttatccactcttttacacagtgaaacagagatt 1023       Db 121 CGACGAGCTCTTCATATTAAAGCTTAAACCTTATCCCATCTCTACACACAGTGAAGGAGTAGT 180
Oy	1024 agagagtgtttcatttcatttttagagctcaagccaaaagagagactgttttttaacagg 1083     Db 181 AGCGGGTGTTCATTATTTTGAAGTCTACGCGCAAAGAGGAGAACGTGTGTGTTTAAACAGGG 240

Query Match	Best Local Similarity	Score	DB	Length	409;
Matches	405; Conservative	0;	Mismatches	4;	Indels 0; Gaps 0;

[illegible]

Standard Commission Act

[illegible]

ACCESSION	R59283	
VERSION	R59283.1	GI:829978
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
1 (bases 1 to 446)  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homidae; Homo.

**AUTHORS** Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasaks, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

**TITLE** The WashU-Merck EST Project

**JOURNAL** Unpublished (1995)

**COMMENT** Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 451  
High quality sequence strops: 315 Source: IMAGE Consortium, LBNL  
This clone is available royalty-free through LBNL; contact the  
IMAGE Consortium (info@image.lbnl.gov) for further information.  
Insert length: 451 Std Error: 0.00  
Seq primer: M13Rpr1  
High quality sequence stop: 315.  
Location/Qualifiers  
1..446

```

/organism="Homo sapiens"
/db_xref="GDB:414339"
/db_xref="taxon:9606"
/clone="IMAGE:41798"
/clone_1ib="Scores infant brain INIB
/sex="female"
/dev_stage="73 days post natal"

```

/lab host="DH10B (ampicillin resistant)"  
 /noise="Organ: whole brain; Vector: lafm1d BA; Site\_1: Not  
 I; Site\_2: Hind III; 1st strand cDNA was primed with a Not  
 I\_oligo(dT) primer [5':  
 AATCGAGAGAAATTCGCGCGCCGACGAAATTTTTTTTTTTTTT 3']:  
 double-stranded cDNA was ligated to Hind III adaptors  
 (Pharmacia), digested with Not I and directionally cloned  
 into the Not I and Hind III sites of the lafm1d BA vector.  
 Library went through one round of normalization. Library  
 constructed by Bento Soares and M.Fátima Bonatodo."

Standard comparison A

Query Match	12.9%	Score 387.2	DB 22	Length 446
Best Local Similarity	98.6%	Pred. No. 2.1e-99		
Matches 411, Conservative	0	Mismatches 4	Indels 2	Gaps 2

```
QY      2594 ttccatgcggaagtcttgaaagtttaatygcccttcaaggacaccagaagcctaagctg   2655
        |||                                     |||||                ||
Db       1 TTTATGCGGAGTTPGAAGAFTTAATGCACCTTCAGGAAACCACAGGGCCCTCACAG-IG    59
```

```
QY      2654 acagtagctaaagcgygaacctctctcatcccttaatgccttgcatattaagatagy 2713
        |||
Db       60 ACAGTAGCTAAAGCTGGAACCTCCTTCATCCCTTCAATGCCTTTGCTCATTAAGAATATG 119
```

**OY**    2714 acatttaccatcgagggggaaccaagcgttcattgacaaactctagaatacttgaaaaaaty    2773  
|||||  
**Dd**    120 ACAATTACTCATGAGGGGAACAAGACGTTCATTTGACAACTCTGTAAMCCTTGAAAAAATG    179

**DY**    2774   cgcatagtatgcaaatacgcggccagaaacagttagataactacaagaaccacccctcaatcct    2833  
        |||||  
**Db**    180   CGCATGATTGCMAATACGGCCAGAACAGTGAGATACTACAGGAGCCCAACCCTTCATCCTT    239

QY 2334 gatcgagcgtcaagctataagaaaccatcagaatgtccggagttactgacgcaattaaat 20933  
Db 240 GATCGAGCTCAAGCTAATAAGAACCATCAGATGTCGGAGTTATGACGGCAATTAAAT 299			

```
QY      2894  gggattgcacaccgaagaaactttatcacagaatgtcacacagaatttagcgctcgtcgacca 2953
        |||
Db       300  GTCATTGCACACCAGAGAATCTTATCACAATGTGCAACACAGATTAGACCTTGTCGACCA 359
```

Qy	2954 tagacattcaatgcccagaagcaacagtltt-tctccagtcacaaatl tccaanaa    3009             
Db	360 TAGACATTCAATGCCCAAAGCAACAGTTTGTCTCCAGTCACAATNTTTTTCAA    416             

RESULT	5
AU079207	
AU079207	541 bp
LOCUS	mRNA
	FSM
	20-Oct-1999

DEFINITION	ACCESSION	VERSION
AU079207 Sugano mouse brain mncb Mus musculus cdna clone MNCB-3 5', mRNA sequence.	AU079207	GI:6083962

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia.	

REFERENCE  
AUTHORS  
1 (bases 1 to 541)  
Eutneria; Kodentia; Sciurognathi; Muridae; Murinae; Mus.  
Hashimoto, K., Kusuda, J., Toyoda, A., Tanuma, R., Ito, A., Hirata, M.  
Suzuki, Y., Sasaki, M. and Sugano, S.

TITLE	Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method
JOURNAL	Unpublished (1999)
COMMENT	On Jun 22, 1998 this sequence version replaced 01:3247237

**FEATURES**

Contact: Katsuyuki Hashimoto  
Division of Genetic Resources  
National Institute of Infectious Diseases  
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
Email: [khashi@nih.go.jp](mailto:khashi@nih.go.jp)  
URL: <http://www.nih.go.jp/yoken/genbank/>,  
Location/Qualifiers

```

source
1.541
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-3986"
/clone_lib="Sugano mouse brain mncb"
/sex="female"
/dev_stage="adult"
/lab_host="T0P10"
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
was primed with an oligo(dT) primer
[ATGCGCCTTTTTTTTTTTTTTT] double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCTACG] , digested
and cloned into distinct DraIII sites of the pME18S-FL3.
XhoI sites just outside the DraIII sites can be used to
isolate the cDNA insert. Size selection was performed by
exclude fragments <1.5 kb. Library was constructed by
Sugano et al. (University of Tokyo, Institute of Medical
Science). Custom primer for sequencing: 5' end primer
[CTTTCCTCTTAAGCTGCG]"

BASE COUNT      149 a      122 c      125 g      137 t      8 others
ORIGIN

Query Match      11.4%; Score 342.4; DB 69; Length 541;
Best Local Similarity 90.8%; Pred. No. 1.4e-86;
Matches 375; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

QY 2390 caatttgatgctcagatcagatcgtcctgtctcagtcagcaagcgtgtcagctca 2449
1 CAGTTTGGGTTGTGCTACGTGANGTCTGCTTTGTCCAGCTCAGCAAGCTGTTCAGCTT 60
DB 1
QY 2450 ttaaaaaatttatttaagatagcagccactgtgaagagatataaaatctgaattc 2509
61 TTGAAAAAATTATTCAGAGATAGCGGCTCACTGCAGAGATACAAAATTAATTCCTTT 120
DB 61
QY 2510 ttggcctgctcagcctcaagtaaatgtcgtgtagccgctgtgcacacgttgagg 2569
121 TTGCCCATTGCTCATGAGCTACGCAACGTGGCCGTGAGCCGCTTGGCATCAACGTGGAG 180
DB 121
QY 2570 aaatgcgaagcaagtcgaagaagttctatgcgaggttgaagtttaagtaagccctca 2629
181 AAATGCCGAGCAAGTTTAAGAGAGTTCTATGCGAGTTTGAAGCTTGAAGTGATTCCTCC 240
DB 181
QY 2630 aggaaccacagagcctcagagctgacagtagctaaagctgaacctcctcctccctc 2689
241 AGAACCACAGAGGCGATACAGCTGACAGCCAGCAAGCTGGAGCCCTCTCATCCCTTTC 300
DB 241
QY 2690 atgccttgctcatgaagaatagacattacatcattgaagggaacaagacgttcattgac 2749
301 ATGCCCTTGCTTATTAAGATATGACATTTACTCATGAGGGAGCAACAGACGTTTCATTGAC 360
DB 301
QY 2750 aatctagtaaaccttgaaaaaatcgcatgatgccaatacggcgcaggaacagt 2802
361 AATCTAGTAACCTTTGAAAAAATGCGCATGATTGC-AAACACTGCCAAGACAGT 412
DB 361

RESULT 6
AM047433/c 411 bp mRNA EST 18-SEP-1999
LOCUS UI-M-BH1-all-c-07-0-UI.s1 NIH BMAP_M.S2 Mus musculus cDNA clone
DEFINITION UI-M-BH1-all-c-07-0-UI 3', mRNA sequence.
ACCESSION AM047433
VERSION AM047433.1 GI:5907962
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 411)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

```

```

JOURNAL
MEDLINE
97044477
COMMENT
Genome Res. 6 (9), 791-806 (1996)
On Jun 5, 1998 this sequence version replaced gi:3189695.
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m5t@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a Bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized corpus striatum library cDNA library preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLY-A=yes.

FEATURES
source
1.411
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-all-c-07-0-UI"
/clone_lib="NIH BMAP_M.S2"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M.S2 library is a subtracted library derived from
NIH_BMAP_M.S1, which in turn is a subtracted library
derived from a mixture of (normalized) libraries from ten
regions of the mouse brain (cerebellum, brain stems,
olfactory bulbs, hypothalamus, cortex, amygdala, basal
ganglia, pineal gland, striatum, hippocampus). The driver
used for subtraction consisted of a pool of 5,000 clones
from the NIH_BMAP_M.S1 library and a pool of 2,000 clones
obtained from non-normalized and normalized mouse brain
spinal cord libraries.
TAG LIB=NIH BMAP_M.S2
TAG_TISSUE=corpus-striatum
TAG_SEQ=ACGGC"

BASE COUNT      90 a      102 c      90 g      128 t      1 others
ORIGIN

Query Match      10.8%; Score 325.8; DB 64; Length 411;
Best Local Similarity 88.1%; Pred. No. 6.6e-82;
Matches 354; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1954 ccaacctacacacaccactcgggtggtccagtcgacacttcggtgaaggaagtcacatgctc 2013
411 CCACACCTATATATACATATTGCTGCGCTTAGCTGCTCGGTGAAGAGATCATCATGTC 352
DB 411
QY 2014 agttccgcgaacagctggtcgcgggagggcctgatcatcaatgacagatgagttccggagg 2073
351 AGTAGCTGACAAACTGGGCTCAGAGAGAGGCGCTGATTAATCGTCAAGATGAATCTTGAGG 292
DB 351
QY 2074 agaaaagggtgctgctcaaacctaatagtattcaagratattacgaagcgtcacattatg 2133
291 AGAAAAGGTGCTGCTGAAATCTATATATGTTTCAGATTTAGACGCTACCATTAATAGG 232
DB 291
QY 2134 acgcctgttctgctccgcgaggaacatcgaattcgaatcctcctacagaacagga 2193
231 ACCCGTGTGCTGCTGCGCGAGAGAGCAATTCGACTGACCTGCTCCGGAACAGGA 172
DB 231
QY 2194 agggccaactcgttggaacggtggaactttgaacgatgagctccaagaatttagcata 2253
171 AGGCCGACCACTGGAGCAAGTGGAACTTTGAGCTGATGAGCTCGAAAGACTTGCGCTA 112
DB 171

```

QY	2254	ccagatgacaattatgattgggaaccttcacatgcgtgcatagactgtagactaatccta	2313
Db	111	CCAGATGACAACCTTACATTGGGAACCTTTAACTGTGTCCATGAGAGCTGGAGCTAATCTA	52
QY	2314	tcacacatttggaagcagataattttaaagaaccagacgcaa	2355
Db	51	CCACACATTTGGAAGCATATTTTAAAGACCAAAAAAAA	10
RESULT	7		
LOCUS	AA388005	361 bp	EST
DEFINITION		vc85b01.r1 Ko mouse embryo 11.5dpc Mus musculus cDNA clone	23-APR-1997
ACCESSION		IMAGE:789769.5, similar to W1:2065.5 CE00479 CAMP DEPENDENT	
VERSION		PROTEIN KINASE REGULATORY CHAIN + GUANINE NUCLEOTIDE RELEASING	
KEYWORDS		FACTORS ; , mRNA sequence.	
SOURCE	AA388005	GI:2040950	
ORGANISM		EST.	
		house mouse.	
		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS		Eutheria; Rodentia; Sclirognathli; Muridae; Murinae; Mus.	
		1 (bases 1 to 361)	
		Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,	
		Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,	
		Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,	
		Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and	
		Waterston,R.	
TITLE		The Washu-HMT Mouse EST Project	
JOURNAL		Unpublished (1996)	
COMMENT		On Sep 12, 1996 this sequence version replaced gi:1395402.	
		Contact: Marra M/Mouse EST Project	
		Washington University School of MedicineP	
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108	
		Tel: 314 286 1800	
		Fax: 314 286 1810	
		Email: mouseest@watson.wustl.edu	
		This clone is available royalty-free through LLNL ; contact the	
		IMAGE Consortium (info@image.llnl.gov) for further information.	
		MG1:474177	
		High quality sequence stop: 327.	
FEATURES		Location/Qualifiers	
SOURCE		1..361	
		/organism="Mus musculus"	
		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/clone IMAGE:789769"	
		/clone_11b="Ko mouse embryo 11.5dpc"	
		/sex="pooled"	
		/tissue_type="embryo"	
		/dev_stage="11.5dpc"	
		/lab_host="DH10B"	
		/note="Organ: embryo; Vector: pSPORT1; Site_1: SalI;	
		Site_2: NotI; Total RNAs were extracted from 11.5 dpc	
		embryos (excluding placenta and yolk sac). The	
		double-stranded cDNA was synthesized with an oligo (dT)-1	
		primer GAGAGACATGATCTGTGATGGGAGCGGCGCTTTTTTTTTTTTTT	
		3'. The cDNAs were ligated to LL-Sal3a: 5'	
		GCAATTGACGCTGCTATCC 3' and LL-Sal3b: 5'	
		GGATGTGACGTCAAT 3'. The cDNAs were size-selected and	
		amplified by long-range PCR using Ex Taq polymerase for 18	
		cycles. The PCR-amplifiable cDNA mixture went through	
		one round of equalization and was digested with SalI/NotI	
		and cloned into the SalI/NotI sites of the pSPORT1	
		plasmid vector (Life Technologies). The library was	
		constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong	
		Wang."	
BASE COUNT	111 a	88 c	78 g
ORIGIN			84 t

Query Match	Best Local	Match	Score	DB	Length
10.5%	93.5%	0	315.2	DB 33	361
Conservative	0	Mismatches	23	Indels	0
0	0	Gaps	0		
2595	tcctatgcgaggttcgaagtttaatgacccctcaaggaacccacagggcctcaagctga	2654			
Db	7 TCTATGGGAGGTTTGAGAGCTTGATGGATTCCTTCAGAAACCAAGGAGCATACAGGCTCA	66			
Qy	2655 cagtagctaaagctgtgaacctctcctccatccctcattgcttgcattaaagatalga	2714			
Db	67 CACACACCAAGCTGGAGACCCCTCTCATCCCTTTCATGCCCTTGCTTATTAAAGATATGA	126			
Qy	2715 catttctcatatgaggggaacaaagacgttattatgacaaattagtaaaccttgaaaaaatgc	2774			
Db	127 CATTACTATGTAGGGGAAACAGACCTTATTGACATATGCTTAAACCTTGGAAAAATGTC	186			
Qy	2775 gcatagttcacaataagcgcacagaacagtgatagactacagagagccaaaccttcaatctcg	2834			
Db	187 GCATGATTGCCAAACACTGCGCAGAACGTAACGCTACTACAGAGGCCAGCCCTTCAATCCGG	246			
Qy	2835 atgcagctcaagcttaataagaacccatcagatgtccggagtgatgtagcgaatataa	2894			
Db	247 ATGCCCTCAACCTATATAAGAACCATCAGAGATGTCGGGAGTTATGTACGGCAATTAATG	306			
Qy	2895 tgattacaaccagagaactttatcacagatgtcacagatgaagctcg	2946			
Db	307 TGATTGACAAACCAAGGAACTTTATCACAGATGTCAACAGATTAAAGAGCTCG	358			
RESULT	8				
LOCUS	A1003002/c				
DEFINITION	A1003002 317 bp mRNA EST 09-JUN-1998				
	am59g07.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone				
	IMAGE:1683900 3' similar to YR:092565 Q92565 MYELOBLAST KIA0277.				
	; , mRNA sequence.				
ACCESSION	A1003002				
VERSION	A1003002.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
REFERENCE	Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 317)				
	Hillier,D., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,				
	Krizman,L., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,				
	Mattin,J., Moore,B., Schellenberg,K., Stepec,M., Tan,F.,				
	Thaïsing,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.				
TITLE	WashU-NCI human EST Project				
JOURNAL	Unpublished (1997)				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2285771.				
	Contact: Wilson RK				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: est@wustl.edu				
	This clone is available royalty-free through LNL ; contact the				
	IMAGE Consortium (info@image.lnl.gov) for further information.				
	Possible reversed clone: similarity on wrong strand				
	Seg primer: -40ml3 fwd. RT from Amersham				
	High quality sequence stop: 269.				
FEATURES	Location/Qualifiers				
SOURCE	1..317				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:1683900"				
	/clone_1lb="Stratagene schizo brain S11"				
	/sex="male"				
	/tissue_type="schizophrenic brain S-11 frontal lobe"				
	/dev_stage="34 years old"				
	/lab_host="SOLR (kanamycin resistant)"				



```

/bb_xref="taxon:9606"
/clone="IMAGE:1684675"
/clone_lib="Stratagene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Maas
excised. Custom library. AVG insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (unpublished) Stanley Neuropathology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."

```

Query Match	100.9%	Score 268	DB 43	Length 266
Best Local Similarity	180.0%	Pred. No. 1,7e-65		
Matches 266	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy 1892	gccacagaagcgacagcctatccgcygctctgtatgaagttctgtttaagtcattatgcatg	1951		
Db 1	gccacagaagcgacagcctatccgcygctctgtatgaagttctgtttaagtcattatgcatg	60		
Qy 1952	gaccacacccatacaacacattcggatgctcgaagtgagccacttcggatgaaggaagatcatagt	2011		
Db 61	gaccacacccatacaacacattcggatgctcgaagtgagccacttcggatgaaggaagatcatagt	120		
Qy 2012	gcagttgcgcaacaagcttggctcgagggaagccttgatcatatgacaagtatgattccga	2071		
Db 121	gcagttgcgcaacaagcttggctcgagggaagccttgatcatatgacaagtatgattccga	180		
Qy 2072	ggagaaagaatgtgtctcaacctaatagtatgtcttaagatattacaagcctcacattaat	2131		
Db 181	ggagaaagaatgtgtctcaacctaatagtatgtcttaagatattacaagcctcacattaat	240		
Qy 2132	ggaagcgcgtgttgcttgcccgagagc	2159		
Db 241	ggaagcgcgtgttgcttgcccgagagc	268		

RESULT	11
A1842094/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
SOURCE	
KEYWORDS	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
COMMENT	

162 B SBacred comparison

```

A1842094      450 bp    mRNA          EST             14-JUL-1999
UT-M-ANI-a-fg-e-01-u-s1 NIH-BMAP-MRG_N Mus musculus cDNA clone
UT-M-ANI-a-fg-e-01-u-3' , mRNA sequence.
A1842094
A1842094.1   GI:5476307
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (phases 1 to 450)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Dec 20, 1995 this sequence version replaced gi:1133363.
```

On Dec 20, 1995 this sequence version replaced g1.1133363.  
Contact: Chn, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7150, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9690  
Email: mestr@nhi.nih.gov  
The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bona fide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalised basal ganglia library cDNA library Preparation: M.B. Soares lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. Seq primer: M13 Forward  
POLYA-Yes.

FEATURES	SOURCE
Location/Qualifiers	
1. . 450	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/clone="UI-M-AN1-afg-e-01-0-U"	
/clone_1ib="NIH_BMAP_MBG_N"	
/dev_stage="27-32 days"	
/lab_host="DH10B (life technologies)"	
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MBG_N library is a normalized library constructed from mouse basal ganglia. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.	
TAG_LIB=NIH_BMAP_MBG_N	
TAG_TISSUE=basal-ganglia	
TAG_SEQ=GTAC.	
105 a 107 c 107 g 131 t	
BASE COUNT	
ORIGIN	

Query Match	8.7%	Score 263.2	DB 61	Length 450
Best Local Similarity	93.8%	Pred. No. 4.9e-64		
Matches 285	Conservative 0	Mismatches 18	Indels 1	Gaps 1
OY	156	caaatggtatgctg-cctggcaaggctcttgatggttaaatgatactgaagcaacagtc	214	
Db	450	CCAATGGTATGCTGCTCTGCGTGGGCTTTGGATGTTAAAGTGTCTGAGACCAACCAATGC	391	
OY	215	accaggaatgcctgtgacacatctgtaacccctgggaattggagacgaccttggagaatgcattc	274	
Db	390	ACCAGAGATGCGGTGACCAATGCACTGTGGAATTTGGGAATTTGGGACAGCCCTTTGGAGATGCATTC	331	
OY	275	tggacacacacaccccgccatctgaacacatcgttaccagagagagcaagtgaactgctcgca	334	
Db	330	TGGATTAACACCCCTCGCATGCAACCAATCGATTACCGAGGAGACAGCCAGCACTTCCGCCCA	271	
OY	335	tcgaacaaagaagacttaagaagcacatitggagaaatctcgacagatratlgtgcagagactc	394	
Db	270	TTGACACAGAGAGACTTCAAGGCACATATGGAGAAATTAACCGACAGTATATGGCCGGACTTC	211	
OY	395	tgagcccccttatgataatttggaaagggacttaacaatgacagagatcttcctgaagag	454	
Db	210	TGGCTCCCTCCCTATATGTTATGTGGAAGGGCTCTAACAAATGACAGATTCCTGACAAAG	151	
OY	455	agaa 458		
Db	150	AGAA 147		
RESULT 12				
LOCUS	A1608196			
DEFINITION	A1608196	277 bp	mRNA	EST
				21-APR-1999
				vr68b01.y1
				ko mouse embryo 11 5dpc Mus musculus cDNA clone
				IMAGE:789769 5' similar to TR:Q92565 Q92565 MTELOBLAST KIAA0277. ;,
				mRNA sequence.
ACCESSION	A1608196			
VERSION	A1608196.1	GI:4617363		



KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 277)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
The Mashu-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Jun 22, 1998 this sequence version replaced gi:3247167.  
Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

FEATURES  
source This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:474177  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the correct orientation)  
Seq primer: Primer name ambiguous  
High quality sequence stop: 268.  
Location/Qualifiers  
1..277  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:789769"  
/clone\_lib="Ko mouse embryo 11 5dpc"  
/sex="pooled"  
/tissue\_type="embryo"  
/dev\_stage="11.5dpc"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from 11.5 dpc embryos (excluding placenta and yolk sac). The double-stranded cDNA was synthesized with an oligo (dT)-1 primer GAGAGAGACGCTGCTGATGCGACGCGCGCTTTTCTTTTCTTTT 3'. The cDNAs were ligated to L1-Sal3A: 5' GGTATGACGTCGACTATCC 3' and L1-Sal3B: 5' GGTATGACGTCGACTATCC 3'. The cDNAs were size-selected and amplified by long-range PCR using Ex Taq polymerase for 18 cycles. The PCR-amplifiable cDNA mixture went through one round of equalization and was digested with SalI/NotI and cloned into the SalI/NotI sites of the pSPORT1 plasmid vector (Life Technologies). The library was constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong Wang."

BASE COUNT 83 a 72 c 59 g 63 t  
ORIGIN

Query Match 8.0%; Score 239.8; DB 48; Length 277;  
Best Local Similarity 92.0%; Pred. No. 1.9e-57;  
Matches 253; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2595 tctatgcgagattgaagttaatgagcccttaagaagaacagagcctcaagctga 2654  
|||||  
DB 3 tctatgcgagattgaagttaatgagcccttaagaagaacagagcctcaagctga 62  
|||||

QY 2655 cagtaagctagctgaagcctcctcatccctccttagctccttgatcataagaataga 2714  
|||||  
DB 63 CAGCAGCCAGGCTGAGAGCCCTCTCATCCCTTTCATGCCCTTGTATTAAGATATGA 122  
|||||

QY 2715 cattactcatgaggggaagaagcgttcattacaaatcagtaaaacttgaaaaatgc 2774  
|||||  
DB 123 CATTACTCATGAGGGGAACAGACGTTTCATTCACATCTAGTAAACTTTGAAAAATGC 182  
|||||

QY 2775 gcatgattgcaatracgagccagaacagtgagatctacacaggaacccctcatccg 2834  
|||||  
DB 183 GCATGATTGCAGAAACACTGCCAGAACAGTACGGTACTACAGAGCCAGCCCTTCAATCCG 242  
|||||

QY 2835 atgcagctcaagctaatataagaacatcagatgctc 2869  
|||||  
DB 243 ATGCCGCTCAAGCTATATAGAACCATCATCAGATGTC 277  
|||||

RESULT 13  
AI386336 441 bp mRNA EST 27-JAN-1999  
AI386336  
LOCUS  
DEFINITION m97c11.y1 Strataene mouse lung 937302 Mus musculus cDNA clone  
IMAGE:552020 5' similar to TR:Q92565 Q92565 MYELOBLAST KIAA0277. ;  
mRNA sequence.

ACCESSION  
VERSION AI386336 GI:4199799  
KEYWORDS  
SOURCE EST.  
ORGANISM house mouse.  
MUS musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 441)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
The Mashu-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Jun 15, 1998 this sequence version replaced gi:3225813.  
Contact: Marra M/Mashu-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

FEATURES  
source This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:332812  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the correct orientation)  
possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Gldco  
High quality sequence stop: 376.  
Location/Qualifiers  
1..441  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone\_image="IMAGE:552020"  
/clone\_lib="Stratagene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (Kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. 6-8 month old female lung and 1.5 year old male lung source of mRNA. Average insert size: 1.5 kb; UniZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3'-3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT 132 a 112 c 90 g 106 t 1 others  
ORIGIN

Query Match 7.3%; Score 220.4; DB 45; Length 441;  
Best Local Similarity 68.8%; Pred. No. 7.8e-52;  
Matches 302; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 2388 ttcaagtttggtgcctcaactgagatcgtctgtctcaagcagcaagcgtgtcagc 2447  
|||||

```

Db      2  TCCACCTCTGGGTAGCCACAGAGATTCTGCTGAGCCACACTGAGCAAGCGCTGCAGC 61
Oy      2448  tatataaaatttattagatagacccactgtaaagagataaaatctgaattcct 2507
Db      62  TGGTGAATAAAATTCATCAAGATTCGACCTCAAGGCCAACAGAACCTTAACCTCT 121
Oy      2508  ttcttgcatgctcgtggaactgaactgctgtgagccgttgccactaacctg 2567
Db      122  TCTTTGCCATGCTGATGGGTCTCAACACCGCCTCTGTACAGCCGGCTTCACAGACCTGG 181
Oy      2568  agaaactgcgaagcaagltcaagaagtctctacgaggttgaaagtttaagaccct 2627
Db      182  AGAAATCCCTGGGAAGTTTAAGAAGCTTTCTCTGAACCTTGAGAGTTTAACGATCCT 241
Oy      2628  caaggaaccacagagcctcagagcgtgacagtaagcctggaacccctcctcctcc 2687
Db      242  CCTTAACCCAAAGCCCTACAGGAGTGGCTCAAAAAGATGAAACCCACCAAAATCCCT 301
Oy      2688  tcatgacctgtcatlaaagatatagacattactcatgaaagggaacaagaacttcat 2747
Db      302  TCATGCCCTTACTGCTTAAGATGTAACGTTTATTTCATGAAGAAATTAACATTTCTTG 361
Oy      2748  acaactagtaaaacttgaanaaaatgcgcatgattgcaatcacgccaagacagtgag 2807
Db      362  ATAACTCTGTCATATTGAAAAGCTGCACATGATGACAGACACTGTCCGAACCTGAGAC 421
Oy      2808  actacagagaccaacctt 2826
Db      422  ACTGAGGACTTAACCAAGTT 440

```

RESULT 14  
 AI909224  
 LOCUS AI909224 540 bp mRNA EST 01-DEC-1999  
 DEFINITION IL-BT202-250399-011 BT202 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AI909224  
 VERSION AI909224.1 GI:6499904

# KEYWORDS SOURCE ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS HC9P <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188999.  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/seq/gethtml.pl?tl=ILat2-IL-BT202-011.html&t3=250399&t4=1>)  
 Seq primer: puc 18 forward.

## FEATURES source

Location/Qualifiers  
 1..540  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BT202"  
 /sex="female"  
 /dev\_stage="Adult"  
 /note="Organ: Breast; Vector: puc18; Site:1: Sma1; Site:2:  
 Sma1; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No.  
 196,716 - Ludwig Institute for Cancer Research) profiles  
 into the puc 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 BASE COUNT 149 a 127 c 143 g 94 t 27 others  
 ORIGIN

Query Match 7.1%; Score 215.2; DB 62; Length 540;  
 Best Local Similarity 78.0%; Pred. No.26-50;  
 Matches 422; Conservative 0; Mismatches 93; Indels 26; Gaps 14;

```

Oy      1636  caataagagcgagatcacgcctggtctacagtggtgcacatgataggacctcct 1695
Db      3  CACAAAAGAGCGAGTCAACAGCCGCTGTTCTACAGTGGCNGCANGTA--GAGACTCCT 60
Oy      1696  gcaagagatgacgtactatgacctcctcgtgagagattatgatactgatacagata 1755
Db      61  GCAGA-GATGAGCTGCTATGCGCTCTGGAAGANTTTANGTANGTCTNATCAGANG 119
Oy      1756  tggcggagatgctgctccctcaagagcaactgccaaggttggaggaagatgtcaaga 1815
Db      120  AGCCCGGAGANTGCGNGG--CNCAAAAGAACMAACGNCAGANTTGAGAA-AAATGCAAGCA 176
Oy      1816  aatctagaagatgcaaaagcaccacaaga---gcaaaagtctcttgcaaca--- 1870
Db      177  AATCTCAGAAGATGCCAANGCACACACAAAGACACACAGTTCCTTCCNACACAG 236
Oy      1870  -gtcaatacgggagatgagaagaccagaagcgccagctat-cggcgactc---ga 1923
Db      237  TGTCACTATCGGGCGATGAGAGAGCCACAGACGCGCACTATNCCGGGCTCTGAAAGA 296
Oy      1924  tgaagttcgtttaaaggtctatcgc--atggacacacctacacaacatcctgggtgca 1981
Db      297  AGTTCTCGTGTAAAGTCTCATTTGCGANTGACACACACTTACACAAACATTTGGGTGCA 356
Oy      1982  gtggccacttcgtgaaagaaatcatcagtgcaattgcgcgaagaactgtggtccggagg 2041
Db      357  GTGGCAGCTTCGGTGAAGAAGTCAATCATGTCAGTGTGCGGACAAACGNGGCTCGGGAG 416
Oy      2042  ggcctgatatcatgtaagaatgtgctccggagg--agaaaagtgtgtctaaactaatg 2099
Db      417  GGNCGA-CANAGNCAGANAGATTCGAGGAGAAAAAAGTGTGCTCAAAACCTAATG 475
Oy      2100  atgttcaatattcagagcgtccacatgaatgagcgcctgttctgcccggaggcg 2159
Db      476  A-GTTTCATATTTCAGA-GCNCACATTAAATGACGCTG-TTCTTTCGCCGCGAGAGN 532
Oy      2160  a 2160
Db      533  A 533

```

## RESULT 15

AW139646 444 bp mRNA EST 30-OCT-1999  
 LOCUS AW139646  
 DEFINITION UI-H-B11-adr-c-04-0-UI.s1 NCI CGAP-Sub3 Homo sapiens cDNA clone  
 IMAGE:2717575 3', mRNA sequence.

ACCESSION AW139646  
 VERSION AW139646.1 GI:6144364  
 KEYWORDS  
 SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jul 7, 1999 this sequence version replaced gi:5406339.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: [Robert\\_Strausberg@nih.gov](mailto:Robert_Strausberg@nih.gov)

Oligo-dT track not found, Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-GAG clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: [www.bio.lnl.gov/db/tp/image/image.html](http://www.bio.lnl.gov/db/tp/image/image.html)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source

1..444  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2717575"  
/clone\_1ib="NCI CGAP\_Sub3"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pT73b-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NCI CGAP\_Sub3 library is a subtracted library derived from the NCI CGAP\_Sub1 library, which is a subtracted library derived from B1. B1 constitutes a mixture of 21 normalized or subtracted NCI CGAP libraries: NCI CGAP\_Co4, NCI CGAP\_Pr22, NCI CGAP\_Pr28, NCI CGAP\_Co10, NCI CGAP\_Co16, NCI CGAP\_Kid5, NCI CGAP\_Kid12, NCI CGAP\_Kid3, NCI CGAP\_Kid11, NCI CGAP\_Lym2, NCI CGAP\_Br2, NCI CGAP\_Co8, NCI CGAP\_CHL1, NCI CGAP\_Le12, NCI CGAP\_Brn23, NCI CGAP\_Lu5, NCI CGAP\_Lu24, NCI CGAP\_Lu19, NCI CGAP\_GC4, NCI CGAP\_GC6, NCI CGAP\_Brn25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below:  
NCI CGAP\_Kid3 pool 1 LAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids 132376-132391), 1456008-1456775, 1500552-1502855); NCI CGAP\_Kid5 pool 1 LAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI CGAP\_Lu5 pool 1 LAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP\_GC4 pool 1 LAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, LAM 2457-2459, 1475592-1476743); NCI CGAP\_Pr22 pool 1 LAM 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI CGAP\_Co10 pool 1 LAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). Subtraction was performed as previously described (Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_LIB=NCI CGAP\_Lu19  
TAG\_TISSUE=Lung  
TAG\_SEQ=GACACGC

BASE COUNT 101 a 130 c 129 g 84 t  
ORIGIN

Query Match 5.4%; Score 163.6; DB 69; Length 444;  
Best Local Similarity 61.5%; Pred No. 1,2e-35;  
Matches 262; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 2348 acagcaacttgattgtcctgaagagatttaataatcaatttggctgctact 2407  
DB 14 ACGGCCAATTGGAGCTGCTGCTGCGAGCGCTGAGGAGTCAAGCAGTGGGTGCGCAC 73  
QY 2408 gagatcgccttgttctcagctcagaagcggttcagctatttaaaaaatttaag 2467  
DB 74 GAAGTCTCTGCTGCGAGGCCCGGCGCAAGCGCGCACCTGCTCAAGAAATTCAACAG 133  
QY 2468 atagcagccactgtatgaagatatataaataatgaattccttttggccatgctatgga 2527  
DB 134 ATGCGCGCCCTCTGCAAGCAAGCAAGACCTGCTCTTTCTACGCCGTGGTCAATGAGG 193  
QY 2528 ctaagtaacattgctgtgagcgcttgcactaaagcttggagaagaaatgcgaagcaagtc 2587  
DB 194 CTGGACAACGCCGCTGCAAGCCGCTTCACATCACTGGAGAGTGCAGGAAATTC 253

QY 2588 aagaattctatgagcgagatttgaagtttaatgaccctcaaggaaccacagggcctac 2647  
DB 254 AAGAACTTGTTCGCAAAATTTGAGAACCTGACGAGACCCCTGCAAGACCAAAAGCTAC 313  
QY 2648 aggtgacagtagtaagctggaacctcctcctcaacccttcagcttgcctaataa 2707  
DB 314 CGAAGATGATCTCCAAATGAAAGCCCGCTGTGATTCCTTCGTGCGCTCGATCTCAAA 373  
QY 2708 gatagacattactcatatgagggagacaagacgcttcattgacaatctagtaacctgaa 2767  
DB 374 GACCTGACTTTCCTGACAGAGGAGTAAAGACCTGTGATGATGTTGTGAACATCGAG 433  
QY 2768 aaatg 2773  
DB 434 AAGCTG 439

Search completed: April 22, 2000, 11:27:55  
Job time: 9891 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 17:24:28 ; Search time 13270 Seconds

(without alignments)  
-689.422 Million cell updates/sec

Title: US-09-422-999-17

Perfect score: 3013  
Sequence: 1 gatccagcagatgtgtgat.....ccacatttcaaaatgcc 3013

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 821193 seqs, -1518192014 residues

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_p11:\*  
8: gb\_p12:\*  
9: gb\_p13:\*  
10: gb\_p14:\*  
11: gb\_p15:\*  
12: gb\_p16:\*  
13: gb\_p17:\*  
14: gb\_p18:\*  
15: gb\_p19:\*  
16: gb\_p20:\*  
17: gb\_p21:\*  
18: gb\_p22:\*  
19: gb\_p23:\*  
20: gb\_p24:\*  
21: gb\_p25:\*  
22: gb\_p26:\*  
23: gb\_p27:\*  
24: gb\_p28:\*  
25: gb\_p29:\*  
26: gb\_p30:\*  
27: gb\_p31:\*  
28: gb\_p32:\*  
29: gb\_p33:\*  
30: gb\_p34:\*  
31: gb\_p35:\*  
32: gb\_p36:\*  
33: gb\_p37:\*  
34: gb\_p38:\*  
35: gb\_p39:\*  
36: gb\_p40:\*  
37: gb\_p41:\*  
38: gb\_p42:\*  
39: gb\_p43:\*  
40: gb\_p44:\*  
41: gb\_p45:\*  
42: gb\_p46:\*  
43: gb\_p47:\*  
44: gb\_p48:\*  
45: gb\_p49:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_p13:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2957.4	98.2	4278	40	HSU78516
2	2413.6	80.1	2982	12	AF115480
3	1112.8	36.9	2399	12	HSU78517
4	721.6	23.9	3394	40	HSU78168
5	720	23.9	2646	40	AF103905
6	636.6	21.1	3373	12	RN078167
7	393.2	13.1	5900	9	D87467
8	176.2	5.8	59255	44	AC016789
9	141.8	4.7	87767	43	AC014497
10	141.8	4.7	111455	33	AC007549
11	141.8	4.7	189668	33	AC007624
12	132.4	4.4	125681	41	AC008340
13	117.6	3.9	59255	44	AC016789
14	102.6	3.4	209071	35	AC005285
15	99.2	3.3	6568	9	AB002311
16	94	3.1	48096	34	CET2065
17	92.6	3.1	128473	11	AC004241
18	81.2	2.7	152184	32	CET70610
19	71.4	2.4	110000	33	AC005079_3
20	71.4	2.4	224064	33	AC005055
21	64.2	2.1	170128	33	AC005059
22	63.6	2.1	111455	33	AC007549
23	54.2	1.8	22920	34	CET14610
24	52.8	1.8	85181	11	AC004622
25	52.2	1.7	2035	34	DD053884
26	50.6	1.7	5398	7	YSCDC25
27	50.6	1.7	5548	7	SCC25
28	50.6	1.7	19198	8	YSC12142
29	50.6	1.7	125681	41	AC008340
30	46.2	1.5	3087	12	RN024996
31	45.4	1.5	2919	8	AF024633
32	45.4	1.5	3172	8	AF015753
33	44.8	1.5	7218	5	166494
34	44.4	1.5	4278	40	HSU78516
35	43.6	1.4	4062	9	E08769
36	43.6	1.4	4070	5	H0MC36P
37	43.2	1.4	3644	7	YSCDC25A
38	43	1.4	224064	33	AC005055
39	41.8	1.4	3731	34	ANDDC3
40	41.4	1.4	3711	40	AF023130
41	41.4	1.4	3711	40	AF181250
42	40.8	1.4	163415	41	AC010756
43	39.4	1.3	3268	7	SPST6
44	39.4	1.3	29500	7	SPCC1442
45	39.4	1.3	47385	7	SPCC1450

ALIGNMENTS

RESULT 1  
LOCUS HSU78516 4278 bp mRNA  
DEFINITION Homo sapiens CAMP-regulated guanine nucleotide exchange factor II  
ACCESSION U78516  
VERSION U78516.1 GI:4115908



OY	1317	gaactaaagaacatgacccaaga	tgctctggtcgtcggagaaggtcccgacg	ggagaacagag	1376	
Db	1505	GACTTTAAAGAACATGACCAAGATG	CTTTGGTCTGGAGAAAGGTCCACACAGG	AAACAG	1564	
OY	1377	ctcttaacaaagaaactcacg	ccctccgacaaagtatactgtgtacg	aaacactg	1436	
Db	1565	CTTGTATCAAGGAACCTCACAGCCT	CACCAAAAGTATCTGTATGTGCAGGAAC	ACTTG	1624	
OY	1437	aaaaaatttagagca	tttctagaacaatacgcccttgaggaac	ctttaatgaagca	1496	
Db	1625	AAAAAATTTTAGAGCATTTTCTG	AAACAATACGCGCTTGAGGCAC	TTTAATGAAGCA	1684	
OY	1497	cagatctgttttaaatgacttata	tatatgatcagctggtttttatgccaat	ccccagc	1556	
Db	1665	CAGATCTGTTTAAATGACCTTTAT	TATATGATGACCTGTGTTTATG	CCAAATACCAAGC	1744	
OY	1557	tttgcccgacgtggtg	ggccactacacgacagccttcacaa	ggtatacagaacagaga	1616	
Db	1745	TTTGCCCGGACGTGGTGCCCACT	TACACAGCAGCCTTCAACAAG	GTACAGAACAGAGA	1804	
OY	1617	aaatggaattatgccc	tcaacaataaagcgagatccgcgtgt	ttctacagtggtcg	1676	
Db	1805	AAATGGATTATTCCTCTCAACA	AATTAAGCGAGTCATCCGCTGT	CTTACAGTGGGCTG	1864	
OY	1677	ccatgtatgagacccctcc	tcagaagatgacatctatgtgcctcc	cttcggagaagttt	1736	
Db	1865	CCATGTATGGAGACCTCCTCA	AGAGATGACGTATCTATGGCCT	CTTCTGGAGAGTTTT	1924	
OY	1737	atgtatcgtatcacagat	gtgcgcggaatgatgtgcctc	taagaagaaactgccaagat	1766	
Db	1925	ATGTATCTGATCAGATGATG	ATGATGCTGCTCCTCAAGGAG	CAACTGCGCAGT	1984	
OY	1797	tgagaagaattgtcaag	caactctgaagaatgatgcaag	caccacaagaagacacaag	1856	
Db	1985	TGGGAAGATTGTCAACCAAT	TATCTCGAAGATGCCAACCA	AAAGAGCAACAGG	2044	
OY	1857	ttcttttgcaacaat	ttcaatcagggcgatgagag	ggccagaagcgcaagc	ctatccgcg	1916
Db	2045	TTCTTTTGCAACANTTCAAT	TAAATACGGGGATGAGAGAC	CCCAAGAGCGCACTATAC	CGCG	2104
OY	1917	gctctgatgaagttc	gtgtttaaagttctatgtcatg	accaactacacaac	catctggg	1976
Db	2105	GCTCTGATGAAGTCTCTTT	TAAAGTCTATTTGATGAC	CAACACCTTACATTC	CGCG	2164
OY	1977	tgccagtgagccaact	tggtgaagaagatcaag	tcagcttgccgaacg	ctcgggctcg	2036
Db	2165	TGCCAGTGGCCACTTCCG	GTAAAGAAATCTCATGCTCA	GTGGCCGACAAAGCTGG	GGCTTCG	2224
OY	2037	ggaggggacctg	atcatagtaagaatgatgc	gcggagagaaaaggtgtgt	ctcaaccta	2066
Db	2225	GGGAGGGGCTCATATAG	TAGTCAAGATGAGTCCG	AGAGAAAAAGGGTGCTCA	AAACTTA	2284
OY	2097	atgatgtttcaagta	tttaagacgctcaacata	tgaagcgctgtttgct	gtggccgag	2156
Db	2285	ATGATGTTTCAGTATTA	ACACGCTCACCTATATATG	ACGCCGTGTTCGCGCAG	2344	
OY	2157	agcaatctgatact	gaactcccttaccaga	aaaggaagggcccaact	gtttggaacagtg	2216
Db	2345	AGCAATTGATTCAC	TCACTACCTTACCAAC	AGGAAGGCCCACTGTTG	GAACAGTGG	2404
OY	2217	gaactttgaactga	tgatgctccaagaattag	catataccagatgaca	attatgatgtgg	2276
Db	2405	GAACTTTGAACCTATG	ATGAGCTCCAAATTTTAC	ATACACAGATGAC	AAATTTTGATTTGGG	2464
OY	2277	aaacttcaactgtg	tcagatgagcttgaggtta	ttcatcaacatattgga	agagcaataat	2336
Db	2465	AACTTTCAACTGGGTG	CTGATGAGCTGAGCTAT	CTATCATCAACATTTG	GAAGACATATTT	2524
OY	2337	ttaaaagacacag	caactgtgatgttctc	tsaggaatttaagaa	ttcaatcaagttt	2396
Db	2525	TTAAAAACACACAGCA	AACTTGTGATTTGTTCTG	AGAGATTTATTAATCA	AAATTTACGTTTT	2584

QY	2397	gggtcgtcaactgagatctgccttctgttcttcagctcagcaagcgtgttcaagctatataaaa	2456
Db	2585	gggtcgtcaactgagatctgccttctgttcttcagctcagcaagcgtgttcaagctatataaaa	2644
QY	2457	aatttattaaagtaacagcccccctctaaaggagtataaaaactgaattcccttctttggca	2516
Db	2645	aatttattaaagtaacagcccccctctaaaggagtataaaaactgaattcccttctttttggca	2704
QY	2517	tgcgtacaggaactaagtaaacattctgtctgtgagccgcttggcactaacgctgggagaaactgc	2576
.Db	2705	tgcgtacaggaactaagtaaacattctgtctgtgagccgcttggcactaacgctgggagaaactgc	2764
QY	2577	caagcaagttcaagaagttctatctcgagagtttgaagtttaatgagacccttcaagaacc	2636
Db	2765	caagcaagttcaagaagttctatctcgagagtttgaagtttaatgagacccttcaaggaacc	2824
QY	2637	aaaggagccacagagctgtgcagagtaagtaagctgagaaaccctctatccctcttaagcctt	2696
Db	2825	aaaggagccacagagctgtgcagagtaagtaagctgagaaaccctctctcatccctcttaagcctt	2884
QY	2697	tgcctacaaagatatgacatttactcatcagtaaggaggaacaagaagcgttcatgcaaatccag	2756
Db	2885	tgcctacaaagatatgacatttactcatcagtaaggaggaacaagaagcgttcatgcaaatccag	2944
QY	2757	taaaccttggaaaaaaltgcgatgatgtgcaaatgacgagcaagacagtgagatactacagga	2816
Db	2945	taaaccttggaaaaaaltgcgatgatgtgcaaatgacgagcaagacagtgagatactacagga	3004
QY	2817	gccaaaccttaactcctgatgtcagctcgaagctcaataaagaacatgaagatgtccggagtt	2876
Db	3005	gccaaaccttgcacatcctgatgtcagctcgaagctcaataaagaacatgaagatgtccggagtt	3064
QY	2877	atcgtacaggaactaataatgtagttgtagcgaacagagagaaccttatcacagatgtacacagat	2936
Db	3065	atcgtacaggaactaataatgtagttgtagcgaacagagagaacctttatcacagatgtacacagat	3124
QY	2937	tagagcctcgtcgcacacatagacatttccaaltgccaaagcaaacagtttgtctccagttcca	2996
Db	3125	tagagcctcgtcgcacacatagacatttccaaltgccaaagcaaacagtttgtctccagttcca	3184
QY	2997	caatttccaaaaa 3009	
Db	3185	caatttccaaaaa 3197	
RESULT 2			
AF115480	2982 bp	mRNA	ROD 25-JAN-1999
LOCUS	Mus musculus CAM- dependent Rap1 guanine-nucleotide exchange factor		
DEFINITION	mRNA, complete cds.		
ACCESSION	AF115480		
VERSION	AF115480.1 GI:4185566		
KEYWORDS	.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 2982)		
JOURNAL	A brain cAMP-dependent Rap1 guanine-nucleotide exchange factor		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 2982)		
TITLE	Gaudriault,G.E., Takaya,K. and Vale,W.W.		
JOURNAL	Direct Submission		
FEATURES	Submitted (19-DEC-1998) The Clayton Foundation for Peptide Biology,		
source	The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, CA 92037,		
	USA		
	Location/Qualifiers		
	1..2982		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
CDS	/feature_type="Brain"		
	1..2982		

```
/codon_start=1
/product="cAMP-dependent Rap1 guanine-nucleotide exchange
factor"
/protein_id="AAD09132.1"
/db_xref="GI:418567"
/translation="MVAAHAHSQSAEWIAICLDRPLERSSEVDVLIITRLKGVAF
EKFPNLRIOICIGYYENLEKGITLERODIGITNMVAVLASLDKVSPTSDDOAV
TICLIGTAFGEISLDNTPRHATIVYRESSELLRIEDQFKLMEKRYOYMGILAP
PYGVETGSSNDRIPDKEENYPSKKILRAGLILRILISRPNHRIRDKYHLKTYRCC
VGTPELVPMHMQTSCVHSRTOAVGMQVYLLEDVLYNVDERHFQDLYLFRDLDER
EDAPLPTBEKKCEDELDQTMILLQMGDAPMLRNPQQRIVDDLEIYIDEL
HKKLSHLSSTYVKRLAVGLIFESHAGGTVLNFQGEISWYIILKGSVNVYIKYG
VCTLHGEDDEGKLAVNDAPRAASYLRDNCNDELILYHTFGRNPKKTTANLDF
EHODVYLVEKVPAGNRAONQSPQOKTWNAGTPEKILHEHLETRIPESTLNAT
DSVLDPMVHCHVPMPTOLCPALVAHYTHQPSQGEQEMDALNKRKRVILVLIQW
AAMTGDILQEDDVAMAFLEEFYVSDDAHMAAFKQDLPELKVQKQISEDKAPQK
KHHVLIQGFNTGERAKRQPIKGSDEVLKVCIDHTYTTIIVPAASKEVYSVA
DKLQSGEGLIIVKNSGGEKVKLSNDVSFTLTNLGFLACPRQFDSLPFPOE
GPTGVGTGELEMSKDLAOMTYDMELNCHYELILYHTFGRNPKKTTANLDF
LBRNEIQFWMVTEVCLQSOLSKRVOLIKFKITAHCKEYKMLNSFPALVGLSVA
VSRILATMEKLPSKFKFYAEFSLMDPSNHRAYRLTAKEPLIPEPLILIKMT
FTHBKNTEIDNLFNEKMRMIANTATVYRSPNPDAQANKNHODVRSYROL
NVIDNQSTISQMSHRLPRRP"
```

```
BASE COUNT      824 a      685 c      758 g      715 t
ORIGIN
Query Match      80.1%; Score 2413.6; DB 12; Length 2982;
Best Local Similarity 89.6%; Pred. No. 0;
Matches 2650; Conservative 0; Mismatches 254; Indels 54; Gaps 3;
```

```
QY 1 gaccagcgaagatcgtgatataactctcactcgaagaaaggttaaaagcttttaga 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 77 GATCAGAGGAGATGTGACACTAATTTTACGCGCGCTGAAGGAGTTAAAGCTTTAGA 136
QY 61 aattcaccaaatctcccttcatacagatttgctatgtatgtatatacgaatcaga 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137 AATTTCACCCAAACCTCCTTCGTGAGATTGTGTATGCGGTTACTATGAGAACTGAAA 196
QY 121 aggaataacattatcttcgcaaggtgata-tgaaacaactgtatgctg-ctggcag 178
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 AAGGAATCACACTGTGTGCGCAAGGAGATATTGGAACCACTGGTATGCTGTGCTG 256
QY 179 gctcttgatgttaaagtatctgacacgaagtcacgaagatgctgtgacatcgtga 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 GGTCTTGAGATGTTAAAGTGTCTAGACACGACGATCACGAGATGCGGTAAACATCTGCA 316
QY 239 ccctcgggaatttggacggtctttggagagtcacatctggaacaacaccccgcatgca 298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 CTCTGGGAATTGGGACAGCGCTTTGGAGAGTTCATTTGATTAACACCCCTCGCATGCAA 376
QY 299 ccactcgttacagaggaagagcgtgaactgctcgcatcggacgaagaagactccaagc 358
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 CCATCGTTACCGAGGAGAGAGCGCAACTTCTCCGCAATTGAGCAGAGACTTTAAAGGCAC 436
QY 359 tatggagaataatcgaacagtataatggcaggaactcgtgctcctcctatgattatagg 418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 437 TATGGAGAAATATACCGACAGTATGTGGCCGACACTTGTGGCTCTCTATGGGTATGG 496
QY 419 aaacggcctcaacaatgacaagattcctcgaagaagaaacacactcatgtaacctcac 478
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 AAAGGGGCTCTAAACATGACAGATTCCTGCAAGAGAAAT----- 538
QY 479 gtctccttcgtcctgctaaacaacataccaaggtccctcctcagaagaagatcctagact 538
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 -----GTCCCTTCAGAGAAAGATCCTCAGAGCT 564
QY 539 ggaanaatttacgaaatcgcattcctcctcgaagacactacatgtaagaagataaaaa 598
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 GGAATAATTTTACGAATTTGCATTCTCTCGAGCTCCCCACATGATAGAGACAGCAAGAG 624
QY 599 taccacctaagaacatacagaacatgctgtgtggaactgaactggtggaactgagatc 658
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 625 TACCACCTAAAGACATACAGACAAATGCTGTGTGGACTGAGCTGTAACTGGATGATA 684
QY 659 gacgagacacacatgcttcaactcccggaactcaagcgtgttggaactgtgcaagtcctgta 718
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 685 CACACAGACATCTGTGTTCACTCGCGAGCTCAAGCTGTGGCATGTGGCAAGCTTGCTG 744
QY 719 gaagatggtgtcttaacacacgtgagacccgggagcacacttccaaagacttattatc 778
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 745 GAAGATGAGTCTCTCAACCATGTGGACCGAGACGGCATTTCCAAAGAAATATTTATTT 804
QY 779 taccatcttcgtatatacgaacagagatgcccccttgcctactacagaggagaagaag 838
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 805 TATCGATTTCTGTGATACACAGCTGAGATGCCCTTTGCTTACTAGSAAAGAAAG 864
QY 839 gagtctgataagagatcgaagacacacatgctgtgctgtgtcacaagatgggccccagcgc 898
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 865 GAGTGTGATGAAGAACTTCAGACACACCATGTGCTGCTCTCAAGATGGGCCCTGACGCC 924
QY 899 cacatgaagatgactccttcgcaaaccaactgagcagaaggaactgtggaagactagaatc 958
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 925 CACATGAGATGATCCTGGAAACCACTGGCCAGAGACTGTGATGACCTAGAGATT 984
QY 959 atcatgaagagctctctcataatataaagcctatcccatctctacacagtgaaacga 1018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 985 ATCTAGACAGAGCTCCTTCAATATTTAAAGCTTATCCCATCTCTTAACACAGTGAACGG 1044
QY 1019 gagttagcaggtgtctcatttttgatgtcacaagcgaagaggaaggtgtgtttaac 1078
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1045 GAGTTAGCAGGTGTCTCATTTTGTAGTCTCAGGCCAAAGGAGAACTGTGTGTTAAC 1104
QY 1079 cagggggaagaggtacccctcctggtacatattctcaaaagatcagtgatgactc 1138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1105 CAGGGGGAAGAGATCCTCTGGTACATCATCATCTGAAAGATCCGTGATGTACTCATTT 1164
QY 1139 taagcaagaggtgtgtgtcgcacccctgcaatgaaggaagatcggcaagtgacacta 1198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1165 TATGGCAAGGAGTGTGTGTCGACCCCTGCACGAGAGAGATGATTGGCAAGTACTCTTA 1224
QY 1199 gtgaatgattgcccacagagctgcctctacatctcgtcttaagaaataactgcatctta 1258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1225 GTGMAAGATGCTCCAAAGACTGCTCCATATTGTTCTCGGAGAGATTAATGTCACTTCCA 1284
QY 1259 aagatgaagaagagatctcaacccggaactcaaaagggcgttgaaagcgaataacatga 1318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1285 AGAGTGCAGAAAGAACTTCAATGGAATTCGAGGAGAGCTTGAGGCCAATACATGCA 1344
QY 1319 cttaagaacatgaaccaaatgtctgtgtcgtggaaggtcccgacgaggaagagact 1378
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1345 TTTAAAGAACATGACCAAGATGTCTTGTAAGTGTGAGAAAGTCCAGGAGGAGAGACT 1404
QY 1379 tctaatcaagaaactcacagcctcagaanaagtatactgtgtgtcagaagacactgaa 1438
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1405 GCTTAATCAAGAAACTCACAGCCTCAGCAAAAGTATACGTGATGTACAGAAACACTGAA 1464
QY 1439 aaaatttgaacatcttctaaagaacaaacagccttggaggaactttaaagaagaca 1498
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1465 AAGATTTTGAACATTTTCTTGAAGCAATACGCTTTGAGCCATCTTTAAATGAACAACA 1524
QY 1499 gatcctgtttaaagacttataatgtacgtgtgttttaatgccaataaccagctt 1558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1525 GATTCGGTTTAAATGACTTGTGTATGATGCACGTGTGTTTATGCCAAATATCCAGCTT 1584
QY 1559 tgcgggaactgtgtgccaactacacgaacagccttcacaaggtacagaaatgaagaaa 1618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1585 TGCCCTGCTTGTGGCCATATACACGACACAGCTTCTCAAGGATACGACAGAGAGAGA 1644
QY 1619 atggaattatgccttaaaataaagagcgtgactacccgctggtgtctcaatggcgtcc 1678
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1645 ATGGATTTATGCCCTCAACAAACAAAGAGGCGGATCACCTGTTGCAATGGGCGGCC 1704
QY 1679 atgtatgagacctcctcgaagaagatgactatcatgagccttcctggagaggtttat 1738
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1705 ATGTATGGCGATCTCTCCAAAGAAAGATGATGTGGCATGGCTTCTTGAGAGATTCTAT 1764
```



QY	1739	gtacctcgtatccagatgatgtccggatgatctgtccctccctaagagagcaatctgcagattg	1798
Db	1765	gtctctcttatccagatgtgcgcgcgcgagtgatgagcctgtcccttcacgagacagctcccgagactg	1824
QY	1799	gagagaagcttccaagcaaatctccaagaagtgtcaaaagycacccacaaagaagcacaaagttc	1858
Db	1825	gagaaagattgtcagacgcaaatctcagaaagccgaaaaagctccacagaaagacacaaagctg	1884
QY	1859	cttttgcacagcttccaataagcggcga tgaagagcccgaaagcgcgcacatcccgycg	1918
Db	1885	cttttgcacacagttcaaacacagctgcagagagggcccgcaaaacgctcacctattcgtggc	1944
QY	1919	tcctgaataatctcgttttaaggtctcttgcagagacacacactaaacaacattccggtg	1978
Db	1945	tcctatagaggcttttggttcaaggctctacgtctacgcacacacctatcaccatttcgtgtg	2004
QY	1979	ccagctggccacttcggtgtaaggaagtcacacag tgaagttgcgcgaacagctgtgctccgg	2038
Db	2005	ccggtactctccctcgtgcgaaggaagtcacagtcagtcagtcagtcgaacaaactggctcaggg	2064
QY	2039	gaaggcctgatcatagtccaagtatgattcccggaaggaagaagtggtgtctcaacctaat	2098
Db	2065	gaagccctgtatcatcgtcagagatgaactctgagagcgaaanaagtgctctcaaaattcaat	2124
QY	2099	gatcttccaattatgaagcgtccacacatgaagcgtcttgcctgcgcgcgagag	2158
Db	2125	gatgttttcagtattttgacagcgtccaccatttaattgacagcgtcttggcctgcccgagag	2184
QY	2159	caattcagttactcgtactcccttaccagacagaaggaagcccaactgttgaacagtgga	2218
Db	2185	caatttcacactacactgactcccttggccggaaacagaaagcccgaccacacagtggaactggga	2244
QY	2219	acttttgaactgaagagctcccaaaagttttagatcaccagatgaacattatgatttggaa	2278
Db	2245	acatttgaactgatgactctgaaagaccttgccgtaccacagtatcacaccttgcagattggga	2304
QY	2279	ctcttcacactgcgtgcctgaagcctgagactaactatcacacatttgaagagcataattt	2338
Db	2305	ctcttcaactgtgctgcattgagctggagctgactatccacacatttggaaagcatatttt	2364
QY	2339	aaaaagaccacagaacactcttgatattgttccctcgagagagaattaagaattccagtttg	2388
Db	2365	aaaaagaccacagcgcaaaacttggattgttgcctcagagaggttttaaaagaatttcagtttgg	2424
QY	2399	gtctgcactgcagctcgtccttctgtctccagctcaagcaagcgttccagctatlaaaaaa	2458
Db	2425	gtttgcactgcagctcgtccttctgttccagctcagcaaaacgitttcacgcttttgaaaaa	2484
QY	2459	tttattaagaatgacgcccacctgttaaggagatlaaaaaatctgaaattccotttttgcac	2518
Db	2485	ttttatcagaagatgagcgtctcactcgaagagatcaaaaaaaactaaatttcccttttgcacatc	2544
QY	2519	gtcacatgagcaatgaatcatctgtctgtagccgcttgcagcaagcttggagaagaacttgca	2578
Db	2545	gtcatgagcaactcaccaacgctggccgctgtagccgcttggcactaacctggagaaactggccg	2604
QY	2579	agcaagctcaagaagtctctatgcggagtttgaagaatttaatgaaccttcaagaagcac	2638
Db	2605	agcaagctttagagattctctatgaggagatttgaagccttatgatgactcttcagaaacac	2664
QY	2639	agggcctacaagctgcgaagaatgactaagcttggaaacccctctatcccttaatgtccttg	2688
Db	2665	agggcctacagctgcagcagccacagctggaacccctctatcccttctatcccttgcgcttg	2744
QY	2699	ctcacatlaagaatagaacttactcatgaggggaaacaagcttcatitgacatctagt	2758
Db	2725	ctttatttaagattatgacattttactcatgaggggaacaaacgcttattgacaaacttacta	2784
QY	2759	aactttgaaaaaatgctgctgatttgcacaatagcggccagaacagttgaatactacaagagc	2818
Db	2785	aactttgaaaaaaatgctgctgatttgcacaatagcggccagaacagttgacgagagc	2844

QY	2819	caaccctcaatcctgtatgcagcttcacagctataagaacacatcagatgtccgagattat	2878
Db	2845	CAGCGCTTCATCCGGATCGCGCTCAACCTATATAAGAACCATCAGGAGTCCGGAGTTAT	2904
QY	2879	gtacgacgaatgaatgtagtattggacacacagagaagaacttatcacagatgcacacagatta	2938
Db	2905	GTACGGCAATTATTAATGTATTGTACACACAGAGAACTTTATCAGAGATGTCAACAGATTA	2964
QY	2939	gagcctcgtcaccatag	2956
Db	2965	GAGCCTCGAAGGCCATAG	2982
RESULT	3		
LOCUS	RNU78517		
DEFINITION	Rattus norvegicus CAMP-regulated guanine nucleotide exchange factor		
ACCESSION	U78517		
VERSION	U78517.1		
KEYWORDS	GI:4115910		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 2399)		
TITLE	Kawasaki, H., Spingett, G. M., Mochizuki, N., Toki, S., Nakaya, M., Matsuda, M., Housman, D. E. and Graybiel, A. M.		
JOURNAL	A family of CAMP-binding proteins that directly activate rap1		
REFERENCE	Science 282 (5397), 2275-2279 (1998)		
AUTHORS	2 (bases 1 to 2399)		
TITLE	Spingett, G. M., Kawasaki, H., Graybiel, A. M. and Housman, D. E.		
JOURNAL	Direct Submission		
COMMENT	Submitted (18-NOV-1996) Center for Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02135, USA		
FEATURES	On Jan 8, 1999 this sequence version replaced gi:4106961.		
SOURCE	Location/Qualifiers		
gene	1. .2399		
CDS	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	<1. .2399		
	/gene="CAMP-GEFII"		
	<1. .1311		
	/gene="CAMP-GEFII"		
	/note="similar to Homo sapiens CAMP-GEFII"		
	/codon_start=1		
	/product="CAMP-regulated guanine nucleotide exchange factor II"		
	/protein_id="AAD03423.1"		
	/db_xref="gi:4115911"		
	/translation="RVIRLVIQMAAMYGDLLQEDDVAMAFLEEFVYSVSDARMVAARKEQIALEKTYKQISDEKAPKQKHYVLQLOFNFGDEAKROPIRGSDVLPFYVCIDHTPTTIRPVAAVKVEIASVAADKIGSGELLITVKNMGGKIVLAKNDVSVFTLLDINGRIFFACPRGQDSLPFLPQEBPTTCTVTFPELMSKLDAYQMTYDMLFMCYLELELIYHFRGRNFKTTANLDFLRRENIQFWVTEICLSQSKRVQLKCKIKTAAHCKEYKNLSNFFGIWGLSNVAASRLATWEKILPSKRTKYAEESLMDPSRNHKAVALTKAKLEPPLIPMPDLIKDMFTFHEGNKFIINDLVNEEKRMIAANTARIVRRSGPFPNDAAQANKNDHODVRSYVOLNVINDQRTLSQMSHRLPEPRRP"		
BASE COUNT	707 a 517 c 522 g 653 t		
ORIGIN			
Query Match	36.9%;	Score 112.8;	DB 12; Length 2399;
Best Local Similarity	88.5%;	Pred. No. 0;	
Matches 1207; Conservative	0;	Mismatches 157;	Indels 0; Gaps 0;
QY	1646	cgagtcattccgcttgatttcacatgagctgcacatgtagagagactcctgcaagagat	1705
Db	1	CGGGTCATCCGGCTGGTTCTTGCAATGCGGAGCGCATGTATGATGATCTTCCAAAGAAC	60

QY	1706	gagatcatatgagcttccttggaagatttatgatactgtatcaagaatgacccgagtg	1765
QY	1706	gagatcatatgagcttccttggaagatttatgatactgtatcaagaatgacccgagtg	1765
Db	61	gAGGCGCATGCGCTTCTCTGGAGAACTCTATGTGCTGTATCAATGATATGCAGGATG	120
QY	1766	attgtctcccttaagagcaactgccagatcttgagaagaatttcaagcaatcccaag	1825
Db	121	ATGTCGCGCTTTAAGAGACACTGGCTGAGCTGGAGAAAGCTCTCAACAGATTCTCAGAA	180
QY	1826	gatcgaaagcaacacaaagaacacaaagttctcttgcacagttcaatcaacygcgat	1885
Db	181	GAGCGAAAAAGCTCCACAGAAAGAACACAAAGTGCTTTTGACAGCATTTAATACAGGTGAC	240
QY	1886	gagaaagcccaagaagcgccagccatcccgcgctctgatagaagtctgtttaagttcat	1945
Db	241	GAGAGGGCCCGAAGACCGTCACCCGATTCGGCGCTGTATGTAATTTTGTTCAAAGGTCTAC	300
QY	1946	tgcatagcaacacactacacaacattccggtgtccagttgcagttgcaccttcggtgaagaatc	2005
Db	301	TGCATTGACACACCGAATACCACTCCGAGTGCCTGTACCGCCTCCGGTGAAGAAAGTC	360
QY	2006	atcagttcagttgcgcgaagaacttggctcccgggagagccctgatactagtaagaatgat	2065
Db	361	ATCACTGCAAGTAGAGACAAACTTGGCTGTGGGAAGCCGTGATCTTGTCAAGATGAC	420
QY	2066	tccggaagaagaagaagtggtgtgtctaaaccataatgattcagtaattatcgagctcac	2125
Db	421	TCTGGAGAGAAAGAGTGTGTCTTAAACCTTAAGATGTTTCAGTATTTCGACGCTCAC	480
QY	2126	attaaatgaagccgttgttgccttcgcgcgagagcaattcgattacactgactcccttaca	2185
Db	481	ATTATATGAGACCGCTGTGGCTGCGCCGAGACAAATTGCATCTGACTCTCCCTTGCCA	540
QY	2186	gaacagaagaagcccaactgttggagacagtlyggaacttttgaactgatgatgctcccaagat	2245
Db	541	GAAACAGGAGGGCCCGACACACCTGGGAGACAGTGGGAGCCTTGAAGTGTAGCTCGAAGAAC	600
QY	2246	ttagatacccaatgacaatttatgatattggaactcttaactgcgtcatgaagctgag	2305
Db	601	TTCGGTCCAGAGTACAGACAGTATATGATGTGGAACTCTTACACTGTGTGCTCGACGCTGAG	660
QY	2306	ctaactctacacatttggaaagcataattttaaaagacacagcaaaccttgatttg	2365
Db	661	CTAATCTTCCACACATTTTGGAGGCAATATTTTAAABAGACACAGCAAACTTGGATTGG	720
QY	2366	ttcctgagagatitaaatgaatcagtttttggctgacagagatctgccttggctt	2425
Db	721	TTCCTGAGAGATTTAATTAATGAATTCAGTTTGGGTTGTACAGATTCGCCCTTGTGTTCC	780
QY	2426	cagctcagcaagcggtgttcagctactataaaaaattatagaatagcgcccaactgtaag	2485
Db	781	CAGCTCACCAAGCCTGTTCAGCTTTTGGAAAAATGTATACAAATATGCGGCTCACGCAAG	840
QY	2486	gagataaaaaatcgaattctccttttgcacatgcataatggagctgaagtaaatgtcgtg	2545
Db	841	GAGTATCAAAAACTTGAATTTCTTCTTTCGGCATGTATGCGGCTCAGTAAGTGGCTGAG	900
QY	2546	agccgcttgcactaaagcttggaagaacttgcacaagcaagttaagaagtctctatgcgag	2605
Db	901	AGCGCGCTGGCATTAACATGGGAGAAACGTGCGAGCAATTTAABAAGTTCTATGCGGAG	960
QY	2606	tttgaagaatttaatgaccccttaagaagaacacagggcctacagcttgacagtaag	2665
Db	961	TTTATAGAGCTTAAATGGATTCCTTCCACAAATATCACAAAGGCTGTACAGGCTGACAGCACTAAA	1020
QY	2666	ctggaacctctctacaccccttcacagcttggcttcattaaagatatgaactttaccat	2725
Db	1021	CTGAGAGCCCCCTCATCCCTTTCATGCGCTTGTATTAAAGATATACATTTACTCAT	1080
QY	2726	gagggagaacaagacttcatcttgacaactcagtaaaacttgaaaaatgcgatattgca	2785
Db	1081	GAGGGGAACAAGACATTTCAATGTGACATCTAGTAATCTTTGAAAAAATGCGATATTGCA	1140
QY	2786	atacggcgcaagaacgttgagatactacagagagccaaccttaactcctgatgacgtctaa	2845

[illegible]

BASE COUNT 721 a 996 c 1042 g 635 t  
ORIGIN MAYVOLKVIDNORELSRLSRELBP"

Query Match 23.9%; Score 721.6; DB 40; Length 3394;  
Best Local Similarity 58.6%; Pred. No. 1.2e-193;  
Matches 1359; Conservative 0; Mismatches 934; Indels 27; Gaps 5;

QY 510 aggtcccttcagagaaagatccttcagagctcggaaatatttcagaatcgtcattctctcc 569  
DB 361 AGGCTTCACAGAGCGGGTCTCAGGGCTGGAGCGAGCTGCATCGCATCTCGGCCA 420  
QY 570 gaacacctcaatgaacagatagaataatcacctaagaatacagacatgctgtg 629  
DB 421 CTCGCCCAAACTCATCCAGACCGGAAGTACCACTTAGGCTCTTCCGCACTGCTGT 480  
QY 630 tgggaactgaactggtgagctgtagatcgacagaaacacatggtgttcactcccgagtc 689  
DB 481 CTGGCGGGAGCTGTGGATGGGATCTTGCCCTGGGACTTGGGACTTGGCCATTCGGAGCC 540  
QY 690 aagctgttgcatgtgcaagtcctgttagaagatggtgttccaacacgttggacaag 749  
DB 541 AAGTTGTGGAACTCTCCAGGTGCTCTGTGTAAGGTGCTCCCTTCCTCATGTGAACACG 600  
QY 750 agcacattccaagaatttattatctatcgattcttgatgtagtgaagcaagagatg 809  
DB 601 ACTGGGCTTCACGAGCCAGAT---GCCCAATTCTACCGGTTCCCGGCGCCGAGCCG 657  
QY 810 ccccttgcctactgagagaaagagagtgtagtagagagctccagaaacatgac 869  
DB 658 AGCCCGTGGGAATCATGTGATGGAGAGAGATTGGCCGAG-----CTGTGG 705  
QY 870 tgcgtgtgcaagatgagcccgagcccaatgagatgactctcgcaaacacgtg 929  
DB 706 CCTGTCTCCACGCGGGGCTCGACGCCCTGCTCATCTGGCAGCTTCCAAAGCCCCAG 765  
QY 930 gccagagagctgtagtgaactagagattatctatgaagagcttcttcaatgaagcct 989  
DB 766 GTACAGCCAGGATGTAAGAGCTGGACCTCATCTTGAAGAGCTGCTGCACATCAAGGCTG 825  
QY 990 tatcccatcttctacacaagtgaaacagagtgtagcaagtgcttcatatttgagcttc 1049  
DB 826 TGGCCACACTCTCCAACTGGGTGAAGCGAATTAGCGGCTGTTCGCTTTGAACAC 885  
QY 1050 aagcccaagagagagctggtgtttaaaccagggggaagaaggtacctctgtgatac 1109  
DB 886 ACAGCAAGGACGAGGACCGTGTGTCAGCCAGGCGGACAAAGGCACTTGTGTACATTA 945  
QY 1110 ttctaaagatcaagtgaatgtagtcatcttaacgcaagggtgtgtctgacacctgcatg 1169  
DB 946 TCTGGAAAGGATCTGTCAACGTGTGATCCCATGGCAAGGGGCTGGAGCACCTTCATG 1005  
QY 1170 aaggaatgacttcgcaagtgtagcactagtgatgattcccccagagctgacctatcg 1229  
DB 1006 AGGAGATGATTTTGGACAGCTGGCTGTGTAATGATGACACCCCGGAGCCACCATCA 1065  
QY 1230 tcttaagagaagaataactgccattctttaagagtagaacaagagatlttcaacggatc 1289  
DB 1066 TCTGTGGAGAAATCAACTGTCAATTCCTGTGTGACAAAGCAGGACTTCAACCGTATA 1125  
QY 1290 taaggagctggaaggaataacagtgtagaactaaagaataatgccaagtgtctgtgctc 1349  
DB 1126 TCAAGGATGTGGAGCAAAACATCGGCTGGAAAGACATGGCAAAAGTGTGTGTGTG 1185  
QY 1350 tggagaaggtcccaagaggaacagagcttctaataagaaactcacagcctcaagcaaa 1409  
DB 1186 TGGAGAGAGCTCTCAGGGGCGCGGCTTCCCGACCCCAACCCAGGCAAGGAACCG-- 1244  
QY 1410 agtatctgtatgtcaagaacacctgaaaaaatttagagcaatttctagaacaatc 1469  
DB 1244 -GTATACAGTATGTCTGGACTCCAGATAAATCCTTAGAGCTTGTGTGGAGGCGCATGG 1302

QY 1470 gctttagagcaacttaaatgaagaacagatctcgttttaactgaacttataatgtagtc 1529  
DB 1303 GACTAGATTCCAGTCTCTATGACCCAAAAGAAATTCCTCATGAGGACTTCTCCTGACCC 1362  
QY 1530 actgtgttttattagccaataatccagcttggccggagcagtggtggccactacaagcag 1589  
DB 1363 ACAGAGCTTTCATGCGCCAGCCCAACTCTGCGTCCCTTCGACACACTTCCATGTGG 1422  
QY 1590 agccttcaagaagtacagaagaagagaataatgattatgtcccttaaatatgaagggcag 1649  
DB 1423 AGCTCGGGGTGGCAGCAGAGAGCGCACCTTACGTGTGCAACAGAGGAGCAGAGA 1482  
QY 1650 tcatccgcctgttcttaagtggtgctgcatgtagtagagactctcctgaagaagatgacg 1709  
DB 1483 TCTTCGGGCTGTGTCAGCCAGTGGGTGGCTGTATGGCTCATGCTCCACACTACCTG 1542  
QY 1710 tatctatgacctctgtagagagtttattgtagtcatgtagatgtagcgcagatg 1769  
DB 1543 TGGCCACAGCTTCTCCCAAACTCTCAGACCTGTGTGGCAGGAGACACCCGACTCAGA 1602  
QY 1770 ctgacctaaagagagaaactgccaagattggag-----aagattgtcaagaatctcag 1823  
DB 1603 ACTGCTGAGGGAGCAGTGGCCAGAGAGCGGCATGCGCACAGGTTGAGATGGCTGTG 1662  
QY 1824 aagatgcaaaagcaaccacaaagaagcaaggttcttltgcaacagttcaatagcgcg 1883  
DB 1663 GGAATGCAATCTCTCAGATGAAGAGCCCGGAACCTTGCTTTGGCTCCCCACACGAGAG 1722  
QY 1884 atgagaagcccaagaagcgccagacctacacgcygctctgataagattctgtttaagctc 1943  
DB 1723 AGCCCTCTCTGCGCAGCAGCTGTGCTGACATCCAAAGTTGGGATAAAGTCCCTATGACATCT 1782  
QY 1944 attgatatgacacacccatacaacaaatctgggtgccaagtgtcgaatctgggtgaagaag 2003  
DB 1783 GCCGCGCAGACCCACCACTGATTGTAACCTGACGCTCGTGTACACCTCCCTGTGAGAGG 1842  
QY 2004 tcatcagtgcaagtgccgacaacagctgggtctccgggaagggcgtatcatagtaagatga 2063  
DB 1843 TGATGGCAGCGTGGCCCAAGAGAGATGGCTGACCAAGGGCAGAGTGTGTGAAGGTCA 1902  
QY 2064 gtccggaaggaagaaagtggtgctcaaacctaaatgattgttcaagatlttaagcgtca 2123  
DB 1903 ATTTGCAAGGATGATCCATTTGGGCTGCGACGAGATGCCCGTGTGGCCACATCTGTGG 1962  
QY 2124 ccaatattggaagcctgttcttgcccggaagcaatctgattcaactgactcccttac 2183  
DB 1963 GGCTCAATGAGGATCTCTTGTGTCAACCCCAAGGAAGTGCATGATGTATCCACACC 2022  
QY 2184 cagaacaggaagggccaactgttgaacagtgaggaaactttagaactgtagtgcacaag 2243  
DB 2023 CTGACCCAGCTGGGCGCCACTGTGGCTCTGCTGAGGGGCTGGACCTGGTGAAGCCAAAG 2082  
QY 2244 atttagtaccagaatgaatltatgatttgggaacttccaactgctgtgataagctg 2303  
DB 2083 ACCTGGCAGGCCAGTGTGACGACACGACGTGAAGCTCTTCAACAGTATCCACACAGGTGG 2142  
QY 2304 agctaatctatacaacatttggaaagcaatatt---taaaagaaccaagacaacttgg 2360  
DB 2143 AACTCATCACTATGTGTGGGCCCCAGCATCTGCGGAGTGTACACCCCAACCTGG 2202  
QY 2361 attgtctcgaaggaattaaatgaattcagtttggctgtagtgcataagatgagctt 2420  
DB 2203 AGCGTTATGCGCGCTTTCANATGACTGCACTGATGGGTGGCCACGACGTTGTGTCT 2262  
QY 2421 gtctcagctcaagcaaggtgttcaagctatlaaaaaattatlaagaatagcaagccact 2480  
DB 2263 GCCCGGTGCCGGGCCCCGGGCGCAGCTGTCAAAAATTCATTAAAGTGGCGGCCACCC 2322  
QY 2481 gtaagagatataaaaatctgtaattcccttlttgccatctgtagtgaagtaagatg 2540  
DB 2323 TCAAGGACACAGAAATGTCAATTCCTTGTGGCGGTATGTTGGCTCGACGAACCTCGC 2382  
QY 2541 ctgttagcgcgttggcctaagcgttggagaaacatcgcaagaagttcaagaagtctcatg 2600

Db 2383 CCATGACGGCGCTAGCCACACACTGGAGCGGCTGCTCAACAAGATCCGGAAGCTGTACT 2442  
QY 2601 cggagtttgaaagtttaatgagcccttaaggaaccaagaagggcttaagagctgagcaatgag 2660  
Db 2443 CCGCCCTCGAGAGGCGCTGCTGATCCCTCATGTGACACCGGGGTATACGACTGCGCCCTCG 2502  
QY 2661 ctaagctggaacctctctcatccctctcatgagcttgctgcttaataagatagatcat 2720  
Db 2503 CCAAGCTCTCCCTCTGCTGCTCATCCCTTCATGCCCCCTCTTCTCAAGACATGACCTTCA 2562  
QY 2721 ctcatgagggggaacaagacgttcatatgacaacttaagtaacttggaaaaaagtgagcatga 2780  
Db 2563 TTCATGAGGGAAGAACACACACACTAGTGAGAAATCATCACTTGAGAAATGAGAAATGA 2622  
QY 2781 ttgcaaatagcgccgaagcaagtgagatactactacaagagcca 2820  
Db 2623 TGGCCAGAGCCGCGGATGCTGCACCACTGCCGAAGCCA 2662

RESULT 5  
AF103905 mRNA PRI 06-DEC-1998  
LOCUS AF103905 2646 bp  
DEFINITION Homo sapiens Rapi guanine-nucleotide exchange factor mRNA, complete cds.  
ACCESSION AF103905  
VERSION AF103905.1 GI:3978530  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2646)  
AUTHORS de Rooij,J., Zwartkruis,F.J.T., Verheijen,M.H., Cool,R.H.,  
Nijman,S.M., Wittinghofer,A. and Bos,J.L.  
TITLE Epac is a Rapi guanine-nucleotide-exchange factor directly  
activated by cyclic AMP  
JOURNAL Nature 396 (6710), 474-477 (1998)  
MEDLINE 99068645  
AUTHORS de Rooij,J., Zwartkruis,F.J.T. and Bos,J.L.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-1998) Laboratory for Physiological Chemistry,  
Utrecht University, Universiteitsweg 100, Utrecht 3584 CG, The  
Netherlands

FEATURES  
source  
Location/Qualifiers  
1..2646  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="12q13"  
/tissue\_type="muscle"  
1..2646  
/note="contains CAMP binding domain; Epac"  
/codon\_start=1  
/product="Rapi guanine-nucleotide exchange factor"  
/protein\_id="AAC83381.1"  
/db\_xref="GI:3978531"  
/translation="MTLRMHRPRSCYQLLEHQRSCIGIKMTPLINSESLDPS  
ESLSQASTERVLAGROLHRHLATCPENLRDRKYLRLRQCCSRRLVDGLIAGL  
GHSRSQVVGICQVLLDEBALCHVKHDMWAQDDAFYRPPGEPVTHHEEELIA  
EAVALLSQRPDALLVYALRPGORDELDLIFELAIKVAHLINSVKLELAV  
LLEPHSKAGTIVLESGDGKTSWYTIWKGSVNVYTHGKGLVTLHFGDDGOLATIND  
APRAATILLREDNCHLRVDKOPNRIKRYEAKTRLEHGVVLYLERASGAGPS  
RPPPGANKRTVMSGPEKLTLELLLEAMGDSADHPTEFLSDPLTRHVPMSQL  
CAALLHIFHEPAGSGEQEKSTVYCNKRQDILRVQWALYCSMLATPVPATSLQK  
LSDLVGDRTLNMLEPOMERRCHLECGCNASAPOMKARMLPYMLNDEPLEGS  
SCAIQVGDKPYDICRPDSVLTFLQVTSVEVMAALAEQDGMTRKGVIVYVNSG  
DAIGLOPDARGVATSLGINRLFLVNPQEAHETLPPDOLGPTVYGSABGDIYASDL  
AGQUTBDMSLFNSIHQVELIHYEPQHLRDTVTLANLEFMRNHELTQWATELCL  
CPVGGPQQLLRKFIKLAHLKQKNLNSFAVWFGLSNFASRLNATWERLPHKVKR  
LYSALERLDPSSNNHRYRLALAKLSPPVLPMPRLKIDMTFTHEGNHLVERLLNFE  
KRMMAAARMLHHRSHNPVPLSPRSRVSHLEDSSQVARISTCEQSLSTRSPAST

BASE COUNT 543 a 825 c 781 g 497 t  
ORIGIN  
MAVVOQLKVIDNORELSRLRELPEP\*  
Query Match 23.9%; Score 720; DB 40; Length 2646;  
Best Local Similarity 58.5%; Pred. No. 3.4e-193;  
Matches 1358; Conservative 0; Mismatches 935; Indels 27; Gaps 5;  
QY 510 aggtcccttcagaagaatcctcaagagcttgaanaaalttaagaaatgcattctctc 569  
Db 146 AGGCTTCACAGAGCGGGTGTCTCAGGGCTGGAGCGACAGCTGCATCCGACATTCCTGGCCA 205  
QY 570 gaagcaccatcagataaagaagatagaanaataccactaaagaacatcagaatgtgtg 629  
Db 206 CTTGCCAATAACTCACCCGAGCCGGAATGACCACTTGAAGCTCTTATGCGAGTGTCTGT 265  
QY 630 tgggaactggaactgtggaactgagatgatcgacgagacacatgtgttcaactccgagtc 689  
Db 266 CTGGCGGGAGCTGTGTGATGGGATCTTGCCCTGGGACTTGGGGTCCATTCGCCGAGCC 325  
QY 690 aagctgttgcatgtggaagtcgttgaagaatgtgttcaacacagtggaacagg 749  
Db 326 AAGTTGTGGGAATCTGCACAGTCTGTGATGAAGTGCCTTCCTGCATGTGAACACG 385  
QY 750 agcacaattccaagaatttattatctatcgtattctgtatgtatgacgagagatg 809  
Db 386 ACTGGGCTTCCAGAGCCAGAT--GCCAATTACCGGTTCCCGGGCCCGGACCCG 442  
QY 810 ccccttgcctactgaggaagaagaagagatgtatgagagatccagacacatgc 869  
Db 443 AGCCCGGGAGACTCATGAGATGAGAGAGAGAGTGGCCGAG-----CTGTGG 490  
QY 870 tctgtgttcaagatgtggcccgagcccaatgagatgaccttgcgaacaacctg 929  
Db 491 CCTGTCTTCCAGCGGGGCTCGAGCCCTGCTCATCTGTGGACTTGAAGACCCCGAG 550  
QY 930 gccaagagactgtgagtgacctgagatatactataggaagctctctcattaaagct 989  
Db 551 GTACGCGACGGATGAGAGAGCTGACCTCATTTTGAAGAGTGGCTGCACATCAAGGCTG 610  
QY 990 tatccatcttctaccacaagtgaacagagatgaagatgttctcatctttgagtc 1049  
Db 611 TGGCCACACTCTCAACTGCGGTGAACGAGAAATTACGGCTGTTCGCTTTGAACCC 670  
QY 1050 agcccaagagagactgtgttgaaccagggggaagaaggtacctctgtatcata 1109  
Db 671 ACAGCAAGGAGGACCGGTGTGTCAGCCAGGCGACAAGGCACTTGTGTACATTA 730  
QY 1110 ttctaaagatcagatgagatgtatctttaaagcggaaggtgtgtctgaccctgcatg 1169  
Db 731 TCTGGAAGGATCTGTCAACAGTGTACCCATGCGCAAGGGCTGGAGACACCTGCATG 790  
QY 1170 aaggaagatgacttcgcaagttgagcactagttaatgatccccaagagctgacctatcg 1229  
Db 791 AGGAGATGATTTTGGACAGCTGGCTGTGTGATGATGACACCCGGGAGGACACCATCA 850  
QY 1230 tcttaagagaagataacttccattcttaagagtagaagaagatcttaacagatcc 1289  
Db 851 TCTCTGGAGAGCAACACTGCTCATCTTCTCGTGTGACCAAGCAGCACTCAACCTATCA 910  
QY 1290 taaggagacttgagggggaatagtcagacttaagaagaatgaagaaagtgtctgtggc 1349  
Db 911 TCAAGGATGGGAGGCAAAAGCCATCGGCTGGAAGAAATGCAAAAGTGCTGTGTC 970  
QY 1350 tgaagaaggtcccaagaaggaacagaagcttcaatcaagaagaaccaagctcagaacaa 1409  
Db 971 TGGAGAGAGCTCTCAAGGCGCGGCTTCGCCAGCCCAACCCAGCAGAGAACG-- 1029  
QY 1410 agtatactgtatgtcagaagaacacctgaanaaattttagagcatctctagaacaatcc 1469  
Db 1029 -GTATACATGATGCTGTGCACCCCGAGAGAAATCTAGACTTCTGTGTGAGGCGCATGG 1087

QY 1470 gcttgaagcaactttaaatagaacacagatctgltttaaataatgacttataatgac 1529  
 DB 1088 GACCAGATTCAGTGCATGACGACCAAGACATTCCTCAGCGACTTCCTCAGCC 1147  
 QY 1530 acgtgttttataagccaataaccagcttggccgagcactgtgtgcccactaccagc 1589  
 DB 1148 ACAGGGGCTTCATGAGCCAGCGCCAACTGCGCTCCCTTCGACCACTTCATGTG 1207  
 QY 1590 agccttcaagaagtataagaagaataatgataatgcccacaataagaagcgag 1649  
 DB 1208 AGCCTTGCGGTGGCAGGACGACGAGCGACACCTACGCTGCAACAAGAGCAGCACA 1267  
 QY 1650 tcatccgcctgttctactagtggtgcctgacatgatgagaacctcctgcagaagagatgag 1709  
 DB 1268 TCTTGGGGCTGCTCAGGACGAGTGGGCTGATGGCTTCATGCTCCACACAGACCTCG 1327  
 QY 1710 tatctatgacctcctcgtgagagatttataatgataatgataatgataatgataatg 1769  
 DB 1328 TGGCCACAGCTTCCTCAGAAACTCTCAGACCTGCTGGGAGGAGACCCGACCTCAGCA 1387  
 QY 1770 ctgacctcaagagcaactgcccagattgag-----aagattgtcaagcaaatctcag 1823  
 DB 1388 ACCTGCTGAGGGGAGAGTGCGCAGAGGCGCGGATGCCACAGGTTGAGAAATGGCTGTG 1447  
 QY 1824 aagatgaaagagcacaacaagaagacaagaagttcttcttgcaacagttcaataagcgag 1883  
 DB 1448 GGAATGCAATCTCTCAGATGAAAGGCCCGGACACTTGCTTTGGCTCCCAACAGGAGG 1507  
 QY 1884 atgagagagcccaagaagcgccagctatccgagctcgtatgaagttctgttgaagttct 1943  
 DB 1508 AGCCCTTCTCTGACAGCAGCTGTCATTCATGAGGATTAAGTCCCTATGACATCT 1567  
 QY 1944 atgcatgagcccaactacaacacattcgcgtgagtcagtgccacttgcgtggaagaag 2003  
 DB 1568 GCCGCGACAGCACTCAGTGTGACCTGTGACCTGTGACAGGCTCCGTGAGAGAGG 1627  
 QY 2004 tcatcagtgagtgccgcagacaagcttgagctccgagggagccgcatcagatgaataatga 2063  
 DB 1628 TGAATGAGAGCGTGGCCAGAGAGATGGCTGAGCAAGGAGGAGGCTGCTGTAAGGTGA 1687  
 QY 2064 gtccggaagagaagaaggtgtgtcctcaaaactaatgatttcaagtattacagacgtca 2123  
 DB 1688 ATTCTGAGAGTGAATGACATGAGCTGACGACAGATGCTCCGTGATGGCCACATCTCTG 1747  
 QY 2124 ccaataatgagacgctgttgcctgtgcgcgagagaacattgatactgactccctac 2183  
 DB 1748 GGTGCTATGAGCGCTCTTGTGTGTCACCCACAGAGAGGCGATGATCCACACC 1807  
 QY 2184 cagaacagagaagcccaactgttgaacagtggaacttgaacttgaactgagctccaag 2243  
 DB 1808 CTGACACAGCTGGGGCCCACTGTGGCTGTGCTAGGGGCTGGACCTGTGATGCCAAGG 1867  
 QY 2244 atttaagcaccagatgacatattatgatgtggaacttcaactgctgcgtacgtacgtg 2303  
 DB 1868 ACCTGGAGGCGCAGCTGAGCGAGCAGACGATGAGGCTCTTCAACAGATTCACACAGG 1927  
 QY 2304 agctaactatcaacacatttgaagacataatt---taaaagacacagaacacttgg 2360  
 DB 1928 AGCTGATCCACTATGTGCTGGGCCCCAGCATCTGCGGGATGTCACACCGCAACCTGG 1987  
 QY 2361 atttgctccgagagagatttaagaattcaagtgttgggtgcgtacatgagatctgctt 2420  
 DB 1988 AGGCTTACAGCGCCCTCTTAATGACTGAGTACGTGGGGGCGACACCGAGCTGTGCTCT 2047  
 QY 2421 gtctcagctcagaagcggttccagatataaaaaatttaataagacagacagcact 2480  
 DB 2048 GCCCGGCGCCGGCCCGCCGAGCTGTGTCAGAAAGTTCAATTAAGCTGGCGGCCACAC 2107  
 QY 2481 gtaagagataaaaaactcgaattcctttttgcacatgcatgggactaagtaacttg 2540  
 DB 2108 TCAAGGAGCGAGAAATCTCAATTCTCTTCTCTTCTCTGATGTTGGCTCAGCAACTCGG 2167  
 QY 2541 ctgtgagcgcttgagcactaagctgggagaacatgcacaagcaagtccaagaagtctatg 2600

DB 2168 CCATGACCGCGCTACCGCACACCTGGAGCGGCTGCCCAAAAGTCCGAAAGCTGACT 2227  
 QY 2601 cggagttggaagtttaataatgaaccttaagaagaacaaagggcttacaagctgacagtag 2660  
 DB 2228 CCGCCCTGAGAGGCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2287  
 QY 2661 ctaagctggaacctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2720  
 DB 2288 CCAAGCTCTCCCT 2347  
 QY 2721 ctcaatgagggagaagaagacgttccatgacaatctagaacttgaanaaaatgacatga 2780  
 DB 2348 TTCATGAGAGGAAGAACACACACTAGTGAAGATTCATCACTTTGAGAGATGAGATGA 2407  
 QY 2781 ttgcaataacggccagaacagtgagatactacagagcca 2820  
 DB 2408 TGGCCAGAGCGCGCGATGCTGCACACACTGCCGAGCA 2447

RESULT 6  
 RNU78167  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

3373 bp mRNA  
 Rattus norvegicus CAMP-regulated guanine nucleotide exchange factor  
 I (CAMP-GEFI) mRNA, complete cds.  
 U78167.1 GI:4079656

Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 3373)  
 Kawasaki, H., Springett, G. M., Mochizuki, N., Toki, S., Nakaya, M.,  
 Matsuda, M., Housman, D. E. and Graybiel, A. M.  
 A family of CAM-2 binding proteins that directly activate Rap1  
 Science 282 (5397), 2275-2279 (1998)  
 99074384  
 2 (bases 1 to 3373)  
 Kawasaki, H., Housman, D. E. and Graybiel, A. M.  
 Direct Submision  
 Submitted (14-NOV-1996) Center for Cancer Research/Department of  
 Brain and Cognitive Sciences, Massachusetts Institute of  
 Technology, 77 Massachusetts Avenue, E17-540, Cambridge, MA 02135,  
 USA

FEATURES  
 source  
 gene  
 CDS  
 Location/Qualifiers  
 1..3373  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 1..3373  
 /gene="CAMP-GEFI"  
 197..2851  
 /note="CAMP-GEFI"  
 /note="GEF domain shows highest similarity to RasGEF  
 family; contains CAMP binding domain"  
 /codon\_start=1  
 /product="CAMP-regulated guanine nucleotide exchange  
 factor 1"  
 /protein\_id="AA012739.1"  
 /db\_xref="GI:4079657"  
 /translation="MYLKRMRHPPCCSYLVEFERRRSCIOGLRWPLTNSESLDR  
 VLSQATTEHVKAGKLLYRLHLLATYPTLLRDRKYHLRLRQCCSGRELVDGLALGL  
 GVHSRSQAVGICQVLLDEGALCHYKHMTFODRPAQYRPPGEPQAGTHVDEEELV  
 EAMALLSQRPDALLVYALRPSRPGORDELDLFEELVHIAKAVHLSNKKRELAAV  
 LLFPHSKAGTVLESGDKGSMYIIMKGSVNVYTRKGLVTLTHEDDGGOLAVND  
 APRATIIIRENNCHFLRVDPKOPNRIIRKVEAKTMPLFEHGVVIVLEPTSGAGPS  
 RPPPGKRRRTYVAGSTPEKTELLLEMRPDSRHDPTEFLTDLTHSVENPCTOL  
 FALHLHFEVSEPPAGSGSEDERSTYICNRQOILRVSRWALVSPMLSDVATSF  
 LQKLSIDVSRDTRISNLIRBOYPERRHHLENGCGVSPQTKARNAVPVPHHEPL  
 PSSAGATRVGDKVYADICRDPHSVLTLLHLEVTASVREVMALAHEDHWTKGQVLVYN  
 SAGVGLQDPARAVATSLGNERTLEFVDDQEVLELPPHEGLGPTGSSEMDLVSA  
 KDLAGOLFEHDMNLFNRIHOVELIHVYLGOHILRDVYTALEPFMRPFNFIQVWATE  
 LCLCPVGPRAQLLRKIKLAHLKEQKNLSFFAVWFGLSNALSISLHATWRLRHK

VRKLYSALERLLDPSWMHRYRLALTRLSPPVLPFMDLLKDMTFIHEGNHTLVENLI  
NFERKMMRAVRMLHHRSHSTAPLSRSLRSHVEDSQAISRTSCSQRSLSTRSP  
ASTWAVYOOOLKVINORELSRSLRELEP"  
BASE COUNT 742 a 981 c 973 g 677 t  
ORIGIN

Query Match 21.1%; Score 636.6; DB 12; Length 3373;  
Best Local Similarity 56.4%; Pred. No. 1.6e-169;  
Matches 1329; Conservative 0; Mismatches 984; Indels 42; Gaps 6;

QY 510 aggtcccttcagagaagatcctcagagcttgaaaaatttcaagaatgccaattcctc 569  
DB 342 AGCCACACACAGAGCATGTGCACAAGGCCGGCAACGCTCTGACCGTATCTTTGGCAA 401  
QY 570 gaggacattcagatagatagataaataccaccataagacatacagaagaatgctg 629  
DB 402 CGTACCCCTACCTCATCCGAGACAGAAATACATCTCGACTACATGCGCATGTCTGCT 461  
QY 630 tgggaactgaactggctggactgcatgacgagacacacatgcttcaaccggaactc 689  
DB 462 CTGGCCGGGAGCTAGTGGATGTGGATCTTGCGTGGCTTGGGGTTCACACTCAGGAGCC 521  
QY 690 aagctgttggaatgctggaagctcctgtttagaagaatgcttcaaccagctggaag 749  
DB 522 AAGCTGTGGGCACTGCGCAGAGGTGTGTGATGAGGGTGCCTTTGCCATGTAAACATG 581  
QY 750 agcacatttccaagaacttta---ttattctatcgatttctgatatgagcagagg 806  
DB 582 ACTGGACCTTCCAGACCCAGACGCCCAATTCCTACAGATTCCTCGGACCGGACCCGAC 641  
QY 807 atgccccttgcctactagagagagaagaagagctgcatggaagagcttccagagacaca 866  
DB 642 CTGC-----AGGAACTCATGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 683  
QY 867 tgcctgctgctgacagatgagcccgagccgacacatgagatgataccttcgcaaacac 926  
DB 684 TGGCCCTACTCTCCAGGAGGGGCTGATGCCCTACTACTGTTGACATCCGGAAGTCCC 743  
QY 927 ctggccagagagctgtgatacctcagaagatatctatgagaagcttctcatatataag 986  
DB 744 CGGGTCAGAGGTACATATTAAGAGCTGAGCCATCTTGTGAGAGTCTCATATATCAAG 803  
QY 987 ccttaccatcttctcaccagctgtaacagagagctgagagctgttctcaattttgagt 1046  
DB 804 CGGTGGCTCACCCTTCTCAACTGGTGAAACGGGAACCTAGCTGTCTTTTGAAAC 863  
QY 1047 ctacgacgaagaaggagagctgtgtgtttaaaccagggggaagaagactcctcgttata 1106  
DB 864 CACACAGACAGGAGCACTGTGTGTGACGACGAGGAGCAAGGGTACCTCATGTGTACA 923  
QY 1107 ttattctaaagatcagatgaaatgtagtcaatttaagcagaagggtgtgtctgcacactgc 1166  
DB 924 TTATCTGGAAGGATCTCTCATATGTGTGACCCGCTGCAAGGGGTGTGTGACACGTTGC 983  
QY 1167 atgaagagagatgacttcgcaagattagcactagatgaatgagcccccagagactcctta 1226  
DB 984 ACGAGGAGATGACTTTTGACAGCTGTGTGTGACGAGAGAGAGAGAGAGAGAGAGAGAG 1043  
QY 1227 tgcgttaccgaagaataactgccaattctttaaagatgagtagaagaaggatttcaaccgga 1286  
DB 1044 TCATCCTTCGAGAAATAATACGTCACTTTTGCGGTGTGACAAAGAGAGACTTCAACGCGA 1103  
QY 1287 tccctaaagagctgtgagcgaatacagctcagactaaagaacatgacagaatgtcttg 1346  
DB 1104 TCATTAAGAGATGTGAAACCAAAACATGAGACTGGAAGAACAGGCGCAAAATGTGTGTAG 1163  
QY 1347 tgcctgagagagctcccgacaggaacagagcttcaatcaagaagaactcagagctcagc 1406  
DB 1164 TTTTGGAGAGAACTCTAGAGGTGCTGGCCTTCCCGCCCTCCGACCCCAAGGCAAGAAACC 1223  
QY 1407 aaaaatatactgtgattgctcaggaacacctgaaaaattttagagcatlctttagaacaac 1466  
DB 1407 aaaaatatactgtgattgctcaggaacacctgaaaaattttagagcatlctttagaacaac 1466

DB 1224 GA---TATACGGTAATGCTTGACACCCACAGAAATAATCTAGAACTCTGTTGAGGCTTA 1280  
QY 1467 taagccttgggaacttaataagaagcaacaaatctgtttaaagtcttataatga 1526  
DB 1281 TGAACCGGATTCAGAGCTCATATGACCCACAGACAGAAATCTCATGTGACTTCTCTCTGA 1340  
QY 1527 tgcactgttcttattatgacaaataaccagcttgcgcgagctgtgtgcccactatacag 1586  
DB 1341 CGCACAATGCTTCATTCCTCCGACACAGCTCTTTGGCGCCCTCCGACACACTTTCACAG 1400  
QY 1587 -----cacagccttcaagaagtacagaagaagaataatgattatgcccctcaaca 1637  
DB 1401 TGGAGCCATCAGAGCTCGCGGGGCGAGCGAGAGAGAACACACACTCATCTGCAACA 1460  
QY 1638 ataaagagagagatccagctgcctgttctaagaatggctgacatgataagagactctgc 1697  
DB 1461 AGAGCAGCAGATCTGTGCTGTGTGACCGGGTGGGGCCCTCTACAGCCCATCTCTCC 1520  
QY 1698 aagaagatgacgtatcctatgaccttcctgagagagatttatagtatctgatacagatgag 1757  
DB 1521 GCTCAGATCCCGTGGCCACACAGCTTCCTCCAGAAACTCTAGACCTGTGAGCAGAGATA 1580  
QY 1758 ccgagatgattgctgcccctaagaagaacatgcccagaattggaagaatgttcaagcaaa 1817  
DB 1581 CCCGACTTACCAACTTGTCTAGGGGAACAGTATCCGAGAACGCGACACACAGGTTGG 1640  
QY 1818 tctcagaagatgcaaaagagcccaaaagaagaacaagttctcttgaacaagttcaata 1877  
DB 1641 AGAATGGCTGTGGGAATGATATCTCTCAGACCAAGGCCCGGAATGCATCTGTTGGTTTC 1700  
QY 1878 cggcgatgagagagcccaagaagccgaagct-----atccgagctctgataagttc 1931  
DB 1701 CTAAACATGAGGAACCCCTTCCCAAGCAGCCCTGGGCGCATCCGAGTGGGGCAAAAGTCC 1760  
QY 1932 tgtttaaagctatgcatgagaccacaactacaacaacatltggytgcagltggccact 1991  
DB 1761 CCTATGATATCTGACAGCCCGACACTCGGTGTGACCTGACCTGCGCGGTACAGCTT 1820  
QY 1992 cgttgaagagatcatcagtgacagttgcccagaagcttgggtccggaggggcctgata 2051  
DB 1821 CCGTAGAGGAAGTATGACAGCTTGTGGCCATGAGAACCACTGAGCAAGGGGCAAGTCC 1880  
QY 2052 taagcaagaatgagttccggagagaaagggtgtgtcctcaaaccttaatgatttcaat 2111  
DB 1881 TGGTAAAGGTCAATTCGCGGGTGTATGTGTGTGACAGCAGATGCCCGGCTGTGG 1940  
QY 2112 ttacgacgctacacatlaatgagcgcctgttgccttgcgcgagagaacatlcgaattcac 2171  
DB 1941 CCACATCTCTGGGGCTCAATGAGCGGATCTTGTGTGACCCACAGAAATGACAGAGC 2000  
QY 2172 tgaactcccttaacagaagaagagcccaactgttgaacaagtgtgaaacttttgaactga 2231  
DB 2001 TGAACCCACACCTGAGACACTGGGGCCCACTGTGGGTTCTTGTGAGATGTGAGACTGTAG 2060  
QY 2232 tgaagctcacaagaatttagacataccagatgacaattatgataagagagcttcaacttgc 2291  
DB 2061 TGAAGTGCAGAGACCTGTGGAGGCCAGCTCACAGACATGATGAGACCTTTCACAGCA 2120  
QY 2292 tgcataagctgagactaattcatcacacatttggaaagcataatt---taaaaaagaca 2348  
DB 2121 TCCACAGGTGTGAGAGCTGATGATGTACTGTGGCCCGGACGACCTGCGGAGGTACCA 2180  
QY 2349 cagcaaaacttggaattgttccctgaagagatttaataatcagatttgggtgcgtacatg 2408  
DB 2181 CTGCAAAACCTGGAGCCCTTATGACGAGCTTCAACAGAGCTGAGTCTGTGGTGGCCACAGC 2240  
QY 2409 agatctgacctgttctcagctcagcaagagtggttaagctatataaaataattatlaaga 2468  
DB 2241 AACTCTGTCTCTCCCGGTTCTCTGGCCCGCGGCTTACAGCTACCTCCGGAATTCATCAAC 2300  
QY 2469 taagcagcccaactgaaggagataaaaaatcgaattccttlttgcatcgtcatgtgagc 2528  
DB 2201 TGGAGGCCCACTCAAGGAGACAGAAATCTCAACCTTCTTGTGGGTGCATGTTGGGCC 2360

Standard comparison B

QY	2529	taagtaaacatgctgctgtagccgctctggagcaactaagctggaggaataatgccaagcaagtcca	2588
Db	2361	TCAGCAACTCGGCACATCAACCCGCTGGCCACACACTGGGAGCGTGTGCCCATTAAGTTC	2420
QY	2589	agaagtctatgacggaagttggaagttaatgtagccctccaaggaaccaagagccctaca	2648
Db	2421	GGAACTGTAAGTCGGCCCTGGAAAGTGTGCTGAGACCTTCCTCGGAACCAAGTGTAC	2480
QY	2649	ggctgacagtagtaagctgtgaacctctctcatcccttaccctgaccttgctcataaag	2708
Db	2481	GATTTGGCTCTCCACCAAGGCTCTCTCTCTGTATCATCCCTTTCATGCGCCCTGCTACCAAG	2540
QY	2709	ataaacattactactatagggaaggaagaagtgtaattgagcaatatagtaaacattgaa	2768
Db	2541	ACATATACCTTCATTTATGAGGAAGCAACACACTGTGTGAAAACTCATCAACTTTTGAGA	2600
QY	2769	aaatgcgatgatgattgcaataagcgcagaacagtgatagatactactacagagccaacctcca	2828
Db	2601	AGATCGGATATGGCCGACAGAGCCGTGGGATGCTCCACACATCGCCGGAAGCCAGACACCG	2660
QY	2829	atctcgatgcagctc	2843
Db	2661	CGCCCTCATCACACCAC	2675
RESULT	7	<i>Sequenced confirmation</i>	
Db	DB7467		
LOCUS	DB7467	5900 bp	mRNA
DEFINITION	DB7467	Human mRNA for KIAA0277 gene, complete cds.	PRI
ACCESSION	DB7467.1	GI:1665818	
VERSION	KIAA0277.		
KEYWORDS	Homo sapiens male brain cDNA to mRNA, clone_lib:psPORT 1		
SOURCE	clone:HA6833.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 5900)		
JOURNAL	Nomura, N.		
REFERENCE	Submitted (27-AUG-1996) to the DDBJ/EMBL/Genbank databases. Nobuo		
AUTHORS	Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3		
TITLE	Yana, Kizurazu, Chiba 292, Japan (E-mail:cdna1fo@kazusa.or.jp.		
REFERENCE	URL: http://www.kazusa.or.jp, Tel:0438-52-3930, Fax:0438-52-3931)		
AUTHORS	2 (sites)		
TITLE	Nagase, T., Seki, N., Ishikawa, K. and Nomura, N.		
REFERENCE	Prediction of the coding sequences of unidentified human genes. VI.		
AUTHORS	The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by		
TITLE	analysis of cDNA clones from cell line KG-1 and brain		
REFERENCE	unpublished (1996)		
AUTHORS	3 (sites)		
TITLE	Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y.,		
REFERENCE	Ohara, O., Tanaka, A., Kotani, H., Miyajima, N. and Nomura, N.		
AUTHORS	Prediction of the coding sequences of unidentified human genes. VI.		
TITLE	The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by		
REFERENCE	analysis of cDNA clones from cell line KG-1 and brain		
AUTHORS	DNA Res. 3 (5), 321-329 (1996)		
TITLE	97191544		
JOURNAL	Location/Qualifiers		
REFERENCE	1..5900		
AUTHORS	/organism="Homo sapiens"		
TITLE	/db_xref="taxon:9606"		
REFERENCE	/clone="HA6833"		
AUTHORS	/clone_lib="psPORT 1"		
TITLE	/sex="male"		
REFERENCE	/tissue_type="brain"		
AUTHORS	56..1798		
TITLE	/gene="KIAA0277"		
REFERENCE	56..1798		
AUTHORS	/gene="KIAA0277"		
TITLE	/note="Similar to a C.elegans guanine nucleotide releasing		
REFERENCE	factor homolog (S43368)"		

[illegible]



QY	2155	agggaacttgcattcaactgctccctccatccagaaacaggaagcccaactglttgaaacgt	2214
Db	1012	AGACCTTGCGGACACTTTGAAACCCATTTCGAGAAATVGGAAATCACAGCAAAAGGTCAT	1071
QY	2215	gggaactcttgaaactgtagcgtcccaagaattagcataccagatgaacatttagctg	2274
Db	1072	GAGGATTTTGGGA---ATGAACACTTGGGATCTTGCTCGAATTAATGAATTTTGATTG	1128
QY	2275	ggaactctcaactcgctgcatagcgttgagctaaactctacacatttggaagcataa	2334
Db	1129	GAGCTCTATTCAATTCATTCCAGCAGCAAGGCTGATCTACTTCACGTTCCAGCAGCAAGG	1188
QY	2335	ttttaaaagaccacccaacttgatattgtctccggaggagatttaataatcatt	2394
Db	1189	AAGTGGGGAAACACACTGCAAAATTCACGCCCTTCTGCTCCAGAAATGCATGAAGGTCACCT	1248
QY	2395	ttggctgcctactgtagatctgcgcctctgttctccagctcaagcgcgtgtcagctataa	2454
Db	1249	TTGGGTGGCCACGGAGATTCGTGCTTGACGCCAGCGTGGGCAAGCGAGTGGACGTGTGTAA	1308
QY	2455	aaaaattattaaagatagcagcccaactgtaagaagataaaatcgaattcccttttgc	2514
Db	1309	AAAAATTCATAAAATTTCGGCTCCTACGCAAAAGCCAGAAACCTGAATTCCTTCTTTCG	1368
QY	2515	catgtctatggagactaaagtatacaatctgttgagccgcttggcaactaaqtlggaaanaact	2574
Db	1369	CATTGTATGGGCTTCAACACTCTCTTCTGTCACTGACTGTGCAAGACTGGGAAATAAT	1428
QY	2575	gcccaagcaagttccaagaagttcctatgcggaagtttgaaagttaatgagccctccaagaa	2634
Db	1429	CCCTGGGAAGTTTAAGAAATCTTTCCTGCAACTTGAATTAACAGATCTTCCTCTAA	1488
QY	2635	cccaagggccttaagcgtgacagtaagctaaagctcggaactcctctccatcccttcatgccc	2694
Db	1489	TCACAAAAGCCCTACGAGATGCATTTCAAAAAGATGAAGAACCCACAAAATCCCTTTATGCC	1548
QY	2695	tttgcctaatgaagatagtcattactaacatgaggggaaacaagacgtlcatatgaaactc	2754
Db	1549	CTTATTCCTTAAGATGTATACATTATTTCATGAAGSAAATTAANAATTTTTTGGATTAACCT	1608
QY	2755	agtaaaccttgaaaaaatgcgcatsgtatgcnaaatcagccagacaagatgagatactacag	2814
Db	1609	TGTCAAATTTTGGAAAAGCTGCATATGTATGCCAGACACTGTGCCAACCCTGAGACACTGCG	1668
QY	2815	gagcaaccccttcaatcctcgtatgcagctcaagctaaagaacacataggaatgctcgag	2874
Db	1669	GACTAACCACTT-----TGGTAACTGTCTCCAAAAGAGCATCAAGACTTAAATCTC	1719
QY	2875	ttatgtacggcaattaaatgtgatatgacaacagagaacttatatacagatgtcacag	2934
Db	1720	CTATGTTAAATACACTGTATGTCTATGACAGCCAGCGACTGTGTTGAAGCTCTGCACAG	1779
QY	2935	attagagcctcg 2946	
Db	1780	GATGAGCCTTCG 1791	

## REFERENCE

## AUTHORS

TITLE  
JOURNAL  
COMMENT

REFERENCE  
AUTHORS

2 (bases 1 to 59255)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckery,R., Boguski,V., Boulanger,B.,  
 Brown,A., Castle,L., Collins,S., Collymore,A.,  
 Cooke,P., Dastellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,  
 Ferreira,J., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heath,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,  
 Lebecky,J., Liu,C., Locke,K., MacDonald,P., Margulis,N.,  
 Melwan,P., Moguk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Teifate,S., Tirrell,A., Vassiliou,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A., and Zody,M.

TITLE  
JOURNAL

COMMENT

Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WtBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information -----  
 Center project name: LA108  
 Center clone name: 22\_F\_110  
 -----

\* NOTE: This record contains 76 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1	754:	contig of 754 bp in length
*		gap of unknown length
*	755	1539: contig of 785 bp in length
*		gap of unknown length
*	1540	2284: contig of 745 bp in length
*		gap of unknown length
*	2285	3047: contig of 763 bp in length
*		gap of unknown length
*	3048	3841: contig of 794 bp in length
*		gap of unknown length
*	3842	4639: contig of 798 bp in length
*		gap of unknown length
*	4640	5394: contig of 755 bp in length
*		gap of unknown length
*	5395	6157: contig of 763 bp in length
*		gap of unknown length
*	6158	6908: contig of 751 bp in length
*		gap of unknown length
*	6909	7692: contig of 784 bp in length
*		gap of unknown length
*	7693	8484: contig of 792 bp in length
*		gap of unknown length
*	8485	9262: contig of 778 bp in length
*		gap of unknown length
*	9263	10032: contig of 770 bp in length
*		gap of unknown length
*	10033	10816: contig of 784 bp in length
*		gap of unknown length
*	10817	11601: contig of 785 bp in length
*		gap of unknown length
*	11602	12376: contig of 775 bp in length
*		gap of unknown length



*	12377	13153:	contlg of 777 bp in length
*			gap of unknown length
*	13154	13939:	contlg of 786 bp in length
*			gap of unknown length
*	13940	14734:	contlg of 795 bp in length
*			gap of unknown length
*	14735	15510:	contlg of 776 bp in length
*			gap of unknown length
*	15511	16284:	contlg of 774 bp in length
*			gap of unknown length
*	16285	17070:	contlg of 786 bp in length
*			gap of unknown length
*	17071	17843:	contlg of 773 bp in length
*			gap of unknown length
*	17844	18622:	contlg of 779 bp in length
*			gap of unknown length
*	18623	19407:	contlg of 785 bp in length
*			gap of unknown length
*	19408	20209:	contlg of 802 bp in length
*			gap of unknown length
*	20210	20980:	contlg of 771 bp in length
*			gap of unknown length
*	20981	21754:	contlg of 774 bp in length
*			gap of unknown length
*	21755	22514:	contlg of 760 bp in length
*			gap of unknown length
*	22515	23302:	contlg of 788 bp in length
*			gap of unknown length
*	23303	24106:	contlg of 804 bp in length
*			gap of unknown length
*	24107	24874:	contlg of 768 bp in length
*			gap of unknown length
*	24875	25638:	contlg of 764 bp in length
*			gap of unknown length
*	25639	26379:	contlg of 741 bp in length
*			gap of unknown length
*	26380	27171:	contlg of 792 bp in length
*			gap of unknown length
*	27172	27954:	contlg of 783 bp in length
*			gap of unknown length
*	27955	28728:	contlg of 774 bp in length
*			gap of unknown length
*	28729	29507:	contlg of 779 bp in length
*			gap of unknown length
*	29508	30299:	contlg of 792 bp in length
*			gap of unknown length
*	30300	31100:	contlg of 801 bp in length
*			gap of unknown length
*	31101	31880:	contlg of 780 bp in length
*			gap of unknown length
*	31881	32636:	contlg of 756 bp in length
*			gap of unknown length
*	32637	33420:	contlg of 784 bp in length
*			gap of unknown length
*	33421	34184:	contlg of 764 bp in length
*			gap of unknown length
*	34185	34973:	contlg of 789 bp in length
*			gap of unknown length
*	34974	35763:	contlg of 790 bp in length
*			gap of unknown length
*	35764	36547:	contlg of 784 bp in length
*			gap of unknown length
*	36548	37311:	contlg of 764 bp in length
*			gap of unknown length
*	37312	38101:	contlg of 790 bp in length
*			gap of unknown length
*	38102	38894:	contlg of 793 bp in length
*			gap of unknown length
*	38895	39670:	contlg of 776 bp in length
*			gap of unknown length
*	39671	40443:	contlg of 773 bp in length
*			gap of unknown length
*	40444	41210:	contlg of 767 bp in length

*	41211	41988:	contlg of 778 bp in length
*			gap of unknown length
*	41989	42772:	contlg of 784 bp in length
*			gap of unknown length
*	42773	43548:	contlg of 776 bp in length
*			gap of unknown length
*	43549	44334:	contlg of 786 bp in length
*			gap of unknown length
*	44335	45138:	contlg of 804 bp in length
*			gap of unknown length
*	45139	45923:	contlg of 785 bp in length
*			gap of unknown length
*	45924	46706:	contlg of 783 bp in length
*			gap of unknown length
*	46707	47484:	contlg of 778 bp in length
*			gap of unknown length
*	47485	48254:	contlg of 770 bp in length
*			gap of unknown length
*	48255	49016:	contlg of 762 bp in length
*			gap of unknown length
*	49017	49810:	contlg of 794 bp in length
*			gap of unknown length
*	49811	50603:	contlg of 793 bp in length
*			gap of unknown length
*	50604	51376:	contlg of 773 bp in length
*			gap of unknown length
*	51377	52156:	contlg of 780 bp in length
*			gap of unknown length
*	52157	52936:	contlg of 780 bp in length
*			gap of unknown length
*	52937	53728:	contlg of 792 bp in length
*			gap of unknown length
*	53729	54498:	contlg of 770 bp in length
*			gap of unknown length
*	54499	55296:	contlg of 798 bp in length
*			gap of unknown length
*	55297	56090:	contlg of 794 bp in length
*			gap of unknown length
*	56091	56878:	contlg of 788 bp in length
*			gap of unknown length

Query Match 5.8%; Score 176.2; DB 44; Length 59255;  
Best Local Similarity 98.3%; Pred. No. 1.3e-38;  
Matches 178; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	2829	atccgatgcagctcaagctataagaaccatcagatgtccgagttatgtacgcaat	2888
Db	46315	ATCCTGATGCGACGCTCATTAAGAACATCAGATGTCGCGAGTTATGTACGGCAAT	46256
QY	2889	taaatgtatgtacaacacaggaacttatcatcacagatgtcacacagattagagcctcgc	2948
Db	46255	TAAATGTATGTGACACACAGAGACTTTATCACAGATGTACACAGATTGAGCGCTGTC	46196
QY	2949	gacatagacattccaatgcccacaagaagttgtctcccgatccacaatttcaaaa	3008
Db	46195	GACCATAGACATTTCAATTCACCAAGCAACAGTTGTCTCCATCCACATCTTCAAA	46136
QY	3009	a	3009
Db	46135	A	46135

RESULT 9  
AC014497 87767 bp DNA HTG 16-NOV-1999  
LOCUS AC014497  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.  
ACCESSION AC014497  
VERSION AC014497.1 GI:6436838  
KEYWORDS HTG; HTGS; PHASE2.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster





SOURCE  
ORGANISM  
fruit fly,  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Celnikoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E.,  
Sylrskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Sequencing of Drosophila melanogaster  
Unpublished  
2 (bases 1 to 189668)  
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Celnikoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequelra,A., Sethi,H., Snir,E.,  
Sylrskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
TITLE  
JOURNAL  
COMMENT  
Submitted (21-MAY-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Aug 2, 1999 this sequence version replaced g1:4887256.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 54 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 586: contig of 586 bp in length  
\* 587 566: gap of unknown length  
\* 1823 1822: contig of 1156 bp in length  
\* 1903 1902: gap of unknown length  
\* 2646 2645: contig of 743 bp in length  
\* 2726 2725: gap of unknown length  
\* 3326 3325: contig of 601 bp in length  
\* 3327 3406: gap of unknown length  
\* 3407 4154: contig of 748 bp in length  
\* 4155 4234: gap of unknown length  
\* 4235 5204: contig of 970 bp in length  
\* 5205 5284: gap of unknown length  
\* 5285 6888: contig of 1604 bp in length  
\* 6889 6969: gap of unknown length  
\* 6969 8402: contig of 1434 bp in length  
\* 8403 8483: gap of unknown length  
\* 8483 10183: contig of 1701 bp in length  
\* 10184 10263: gap of unknown length  
\* 10264 11736: contig of 1473 bp in length  
\* 11737 11816: gap of unknown length  
\* 11817 13661: contig of 1845 bp in length  
\* 13662 13741: gap of unknown length  
\* 13742 15560: contig of 1819 bp in length  
\* 15561 15640: gap of unknown length  
\* 15641 17739: contig of 2099 bp in length  
\* 17740 17819: gap of unknown length  
\* 17820 19768: contig of 1945 bp in length  
\* 19769 19848: gap of unknown length  
\* 19849 23150: contig of 3302 bp in length

\* 23151 2330: gap of unknown length  
\* 2331 26360: contig of 3130 bp in length  
\* 26361 26440: gap of unknown length  
\* 26441 32462: contig of 6022 bp in length  
\* 32463 32542: gap of unknown length  
\* 32543 38690: contig of 6147 bp in length  
\* 38690 38770: gap of unknown length  
\* 38770 44144: contig of 5374 bp in length  
\* 44144 44224: gap of unknown length  
\* 44224 53717: contig of 9493 bp in length  
\* 53717 53796: gap of unknown length  
\* 53796 60200: contig of 6404 bp in length  
\* 60200 60280: gap of unknown length  
\* 60281 68813: contig of 8533 bp in length  
\* 68813 68893: gap of unknown length  
\* 68893 82418: contig of 13525 bp in length  
\* 82418 82419: gap of unknown length  
\* 82419 82499: gap of unknown length  
\* 82499 110321: contig of 27823 bp in length  
\* 110321 110401: gap of unknown length  
\* 110402 133615: contig of 23214 bp in length  
\* 133616 133695: gap of unknown length  
\* 133696 168863: contig of 35168 bp in length  
\* 168863 168943: gap of unknown length  
\* 168944 169169: contig of 226 bp in length  
\* 169169 169170: gap of unknown length  
\* 169170 169250: gap of unknown length  
\* 169250 169983: contig of 734 bp in length  
\* 169983 170063: gap of unknown length  
\* 170063 170787: contig of 724 bp in length  
\* 170787 170788: gap of unknown length  
\* 170788 171661: contig of 694 bp in length  
\* 171661 171641: gap of unknown length  
\* 171642 172348: contig of 707 bp in length  
\* 172348 172429: gap of unknown length  
\* 172429 173135: contig of 707 bp in length  
\* 173136 173215: gap of unknown length  
\* 173216 173826: contig of 611 bp in length  
\* 173826 173906: gap of unknown length  
\* 173907 174615: contig of 709 bp in length  
\* 174615 174695: gap of unknown length  
\* 174696 175429: contig of 734 bp in length  
\* 175430 175509: gap of unknown length  
\* 175509 176117: contig of 608 bp in length  
\* 176117 176197: gap of unknown length  
\* 176198 176892: contig of 695 bp in length  
\* 176892 176971: gap of unknown length  
\* 176971 177591: contig of 619 bp in length  
\* 177592 177671: gap of unknown length  
\* 177672 178328: contig of 657 bp in length  
\* 178329 178408: gap of unknown length  
\* 178409 179175: contig of 767 bp in length  
\* 179176 179255: gap of unknown length  
\* 179256 179658: contig of 603 bp in length  
\* 179659 179938: gap of unknown length  
\* 179938 180604: contig of 666 bp in length  
\* 180605 180684: gap of unknown length  
\* 180685 181325: contig of 641 bp in length  
\* 181326 181406: gap of unknown length  
\* 181406 182222: contig of 737 bp in length  
\* 182143 182222: gap of unknown length  
\* 182223 182948: contig of 726 bp in length  
\* 182949 183028: gap of unknown length  
\* 183029 183685: contig of 657 bp in length  
\* 183686 183765: gap of unknown length  
\* 183766 184460: contig of 695 bp in length  
\* 184461 184540: gap of unknown length  
\* 184541 185147: contig of 606 bp in length  
\* 185147 185227: gap of unknown length  
\* 185227 185843: contig of 617 bp in length  
\* 185844 185923: gap of unknown length  
\* 185924 186773: contig of 732 bp in length  
\* 186775: gap of unknown length  
\* 186776 187433: contig of 678 bp in length  
\* 187434 187513: gap of unknown length





REFERENCE 1 (bases 1 to 59255)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone RP11-22F10  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 59255)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lhoczkzy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Tlreli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wymen,D., Ye,W.J., Zimmer,A. and Zody,M.

## Direct Submission

Submitted (05-DEC-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit,A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: I4108

Center clone name: 22\_F\_10

\* NOTE: This record contains 76 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 754: contig of 754 bp in length  
\* gap of unknown length  
\* 755 1539: contig of 785 bp in length  
\* gap of unknown length  
\* 1540 2284: contig of 745 bp in length  
\* gap of unknown length  
\* 2285 3047: contig of 763 bp in length  
\* gap of unknown length  
\* 3048 3841: contig of 794 bp in length  
\* gap of unknown length  
\* 3842 4639: contig of 798 bp in length  
\* gap of unknown length  
\* 4640 5394: contig of 755 bp in length  
\* gap of unknown length  
\* 5395 6157: contig of 763 bp in length  
\* gap of unknown length  
\* 6158 6908: contig of 751 bp in length  
\* gap of unknown length  
\* 6909 7692: contig of 784 bp in length  
\* gap of unknown length  
\* 7693 8484: contig of 792 bp in length  
\* gap of unknown length  
\* 8485 9262: contig of 778 bp in length  
\* gap of unknown length  
\* 9263 10032: contig of 770 bp in length  
\* gap of unknown length  
\* 10033 10816: contig of 784 bp in length  
\* gap of unknown length

\* 10817 11601: contig of 785 bp in length  
\* gap of unknown length  
\* 11602 12376: contig of 775 bp in length  
\* gap of unknown length  
\* 12377 13153: contig of 777 bp in length  
\* gap of unknown length  
\* 13154 13939: contig of 786 bp in length  
\* gap of unknown length  
\* 13940 14734: contig of 795 bp in length  
\* gap of unknown length  
\* 14735 15510: contig of 776 bp in length  
\* gap of unknown length  
\* 15511 16284: contig of 774 bp in length  
\* gap of unknown length  
\* 16285 17070: contig of 786 bp in length  
\* gap of unknown length  
\* 17071 17843: contig of 773 bp in length  
\* gap of unknown length  
\* 17844 18622: contig of 779 bp in length  
\* gap of unknown length  
\* 18623 19407: contig of 785 bp in length  
\* gap of unknown length  
\* 19408 20209: contig of 802 bp in length  
\* gap of unknown length  
\* 20210 20980: contig of 771 bp in length  
\* gap of unknown length  
\* 20981 21754: contig of 774 bp in length  
\* gap of unknown length  
\* 21755 22514: contig of 760 bp in length  
\* gap of unknown length  
\* 22515 23302: contig of 788 bp in length  
\* gap of unknown length  
\* 23303 24106: contig of 804 bp in length  
\* gap of unknown length  
\* 24107 24874: contig of 768 bp in length  
\* gap of unknown length  
\* 24875 25638: contig of 764 bp in length  
\* gap of unknown length  
\* 25639 26379: contig of 741 bp in length  
\* gap of unknown length  
\* 26380 27171: contig of 792 bp in length  
\* gap of unknown length  
\* 27172 27954: contig of 783 bp in length  
\* gap of unknown length  
\* 27955 28728: contig of 774 bp in length  
\* gap of unknown length  
\* 28729 29507: contig of 779 bp in length  
\* gap of unknown length  
\* 29508 30299: contig of 792 bp in length  
\* gap of unknown length  
\* 30300 31100: contig of 801 bp in length  
\* gap of unknown length  
\* 31101 31880: contig of 780 bp in length  
\* gap of unknown length  
\* 31881 32636: contig of 756 bp in length  
\* gap of unknown length  
\* 32637 33420: contig of 784 bp in length  
\* gap of unknown length  
\* 33421 34184: contig of 764 bp in length  
\* gap of unknown length  
\* 34185 34973: contig of 789 bp in length  
\* gap of unknown length  
\* 34974 35763: contig of 790 bp in length  
\* gap of unknown length  
\* 35764 36547: contig of 784 bp in length  
\* gap of unknown length  
\* 36548 37311: contig of 764 bp in length  
\* gap of unknown length  
\* 37312 38101: contig of 790 bp in length  
\* gap of unknown length  
\* 38102 38894: contig of 793 bp in length  
\* gap of unknown length  
\* 38895 39670: contig of 776 bp in length  
\* gap of unknown length





Db 124776 GCTGCCCTCCAAATATACAGCACTGTTCAACGACCTGCAGATTCGATGATCCCTCGCG 124835  
QY 2632 gaacccagagggcctaagcctgaagcgt-----agctaagctggaaacctccctccat 2682  
Db 124836 CAACATGACCAAGTATCGGCACTAGTCCGCCGAACATCTAGTGGCCACACCCCATCAT 124895  
QY 2683 ccccttcagagccttcctcatatgaatagcattactactagaggggaacaagcgt 2742  
Db 124896 CCCGTTCTATCCGATGCTAAGAGAGATCTACCTTTATTCACCTGGGCAATGATACGAG 124955  
QY 2743 catgcaacatagtaacttggaaataatgcgcattgattgcaatagcagaagcgt 2802  
Db 124956 AGTTATGAGCCTAATCACTTTGAGAAAGCTGGCAATCGTGGCAAGAGATGACGCTGCT 125015  
QY 2803 gagatactacagagcaacccctcaatc 2831  
Db 125016 CACGACATGTGCGATTCGCCCATACGATC 125044

RESULT 15  
AB002311  
LOCUS AB002311 6568 bp mRNA PRI 13-FEB-1999  
DEFINITION Human mRNA for KIAA0313 gene, complete cds.  
ACCESSION AB002311  
VERSION AB002311.1 GI:2224566  
KEYWORDS KIAA0313.  
SOURCE Homo sapiens male brain cDNA to mRNA, clone\_1lb:pbluescriptII SK plus clone:HG0186.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 6568)  
Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,  
Miyaajima,N., Kotani,H., Nomura,N. and Ohara,O.  
Direct Submission  
Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo  
Yama, Kizutazu, Chiba 292, Japan (E-mail:cdna@fokekazusa.or.jp,  
URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,  
Fax:+81-438-52-3931)  
2 (sites)  
REFERENCE  
AUTHORS Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,  
Miyaajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
VII. The complete sequences of 100 new cDNA clones from brain which  
can code for large proteins in vitro  
JOURNAL DNA Res. 4 (2), 141-150 (1997)  
MEDLINE 97349984  
FEATURES  
source Location/Qualifiers  
1..6568  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HG0186"  
/clone\_lib="pbluescriptII SK plus"  
/sex="male"  
/tissue\_type="brain"  
63..4562  
/gene="KIAA0313"  
63..4562  
/gene="KIAA0313"  
/codon\_start=1  
/protein\_id="BAA20772.1"  
/db\_xref="GI:2224567"  
/translation="MKPLAIPANHGVMGOQEKSLPADFTKLHLTDSLHPQVTHVSS  
HSGSITSDSGSSSLDIYOATESPAGMDLSGLPETVADSEDDDEEDIRASDPLK  
SRDIYDLEKPDIDRTDDIDLEEFMHOLPAPANMTMSVRELCATMVAVERAG  
TYLNDGEELDSMSYIINGSVYTPDKAKELICGNSFGSPTRDKRYMGVARTKY  
DDCFVCAQDQYCRILNOKVERNMOKVEEBEIVAKVSHRELDRITKGIIVKIGIS  
EKLTMHVEHSHVDPTEFIEDFLTYRFLSSPMVPGKLLMEFNDPLRQKVRVYL  
LWVNHFNDFEDGPMATRFIEFENLIREKMGHLRLNLINACAKAKRRRLMTLTKPS  
REAPLPFILGSEKGFIFVDSVDSKATEAGLKGDOIILEVNGOFENIOISKAY  
ELRNNTHLSTIVKTNLFPVKELLRLRSEKRNAGPHLPKIGDIKKARSYISPLAND

VEOYIGLEKVNKSKANTVGGRRMLKKILDKTRISILPOKRYNDIGIOSDDSIYGL  
ROKRIETALPVGSGTSSNPDLLOSRRHLTDSAPBDLPDYLRFKPDQOSRYTMI  
SKDTTAEVYIOAIRERAVATPDQISLCEVSTPSPVYKORLPLPOLSKLARIDLS  
GRYLLKNNMETETLCSDEQELLRESQSLSDLSVEVATQLSMNFELFRNIEPTE  
YIDDFELRSKTSICANLAREEYINOTEVASIELRNLQKRMKTIHFKIALAC  
RECKNFMSFAIISGLNLAPEVLRFTWELIPMKYKRLFDODLDDLPDRMAKYRN  
LNSONLOPPIIPLPVYIKDPLFLHEGNSKVGVLNKKRRLIAKIEIRHVGMSVN  
MDPALMRTKKKWRSLGSLSGSTNATYVDVQGTGKHKRVRRSFLNAKTIYEDAO  
MARKVOYLSNLELMEDESLQTLSDCEPATNVLKRNPDKKPVKSETPVAPRAGS  
QOKAQSILPQPOQPPPAHKINOGLQVPAVSLYPSRRKVPKDLPPGINSPOLAKIL  
SLSEESLERHKKQAEEDTISNASSQLSPPTSPQSRKGYTLAPGIVNSPOLAKIL  
EISRSISIVNSSEFSDVPLHDERORHESVIVETNGLGRRERTMLPEODYSLS  
YAPMSEGRGLYATATYVISPSTEELSDODGRASLIADAGRSMSWTCSSGSHDNOT  
IOHORSMETLPRGTHHDYGDGDPAGLMASSHHDOIMEFSHSTKYRQMSRLEQA  
QSRASMASSTGVMEDESDGTIKRKGADVSIEAESLSLVSITTEKTPVMPAHAI  
AVASSTTGLIARKEGRREPPPTPEYIGIPTDPEGRSHPARPPPYNVALORSH  
MVARSPTDAPSSVOQPHGHTSSRPVKNPKQMKPNDSPRLAPYOSOGFSREDEDE  
QVSAY"

BASE COUNT 1974 a 1400 c 1463 g 1731 t  
ORIGIN

Query Match 3.3%; Score 99.2; DB 9; Length 6568;  
Best Local Similarity 51.1%; Pred. No. 5.9e-17;  
Matches 289; Conservative 0; Mismatches 268; Indels 9; Gaps 2;

QY 2228 ctgattgagctccaaagatttgatgcatcacaatgacatattatgttggaaactctcaac 2287  
Db 2205 CAGCTCACGACTGTGGAGAGTTGCAACACGCTCTATGCGAAATTTGAACTCTTTCGC 2264  
QY 2288 tggcgagctgagctgagcctaatacaccatttggaaagcctaattttaaagaac 2347  
Db 2265 AACATTGACCTACTGAAATATATGATGATTTATTT--AAACTCAGATCAAAACCAAGC 2331  
QY 2348 aacgaacacttgatttctgctgagagatttaagaaattcagtttggctgctact 2407  
Db 2322 TGTGCAACCTGGAAGAGATTGGAAGAGTATTAACGAGAAATTTTGGTGACATCT 2381  
QY 2408 gagatgctccttggctcagctcagaagcgctgtcagctattataaataattatga 2467  
Db 2382 GAAATTCACAGAAACCAACACGCTGAGAGATGAAGATCAATTAAGCATTTTCATCAAG 2441  
QY 2468 atagagcccccctgtaagagatataaataatcgaattccttggcaccatgcatgga 2527  
Db 2442 ATAGCACTGCACTGAGGAAATGCAAGATTTTAACTCAATGTTTCAATCATCAAGTGGC 2501  
QY 2528 ctaagtaacattgctgtagcgccttgacactacgctgagagaaactgcaagaagcttc 2587  
Db 2502 CTAACCTGCGACCAAGTGGCAAGACTGGAACGACCTGGGAGAAACTCCCAATTAATAC 2561  
QY 2588 aagaagttctatgctgagatttgaagtttaatgagcccttcaaggaaacacagggcctac 2647  
Db 2562 GAAAAACTAATTTCAAGATCTCCAGACCTGTTGATCTCTCCAGAAACATGCGAAATAT 2621  
QY 2648 ag-----gtcgacagtagtcgaagctggaacctctcatcccttcacagccttgc 2701  
Db 2622 CGTAATGTTCAATATGCAAAATCTACAACTCCATTAATCCCTATTCCTCCAGTATC 2681  
QY 2702 attaaagatgacatttaccatgagaggaagaagcgcgttcattcattgacaatctgaatac 2761  
Db 2682 AAAAAGATCTCACCTTCTTCAAGAGAAATACATCAAAAGTAGACGGGCTGTCAT 2741  
QY 2762 ttgaaaaaatgcgcatgattgcaaa 2787  
Db 2742 TTTGGAAGCTAAGGATGATTGCAAA 2767

Search completed: April 22, 2000, 17:50:32  
Job time: 32288 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 16:15:23 ; Search time 351.53 Seconds  
(without alignments)  
1026.159 Million cell updates/sec

Title: US-09-422-999-17  
Perfect score: 3013  
Sequence: 1 gatccagcgagatgtgat.....ccacatttcaaaatgcc 3013

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PCRT059.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.8	1.5	7218	1	US-08-232-463-14
2	35.4	1.2	2652	1	US-08-318-831-1
3	35.2	1.2	5183	4	US-08-870-518-7
4	34.6	1.1	2128	3	US-08-371-377-16
5	34.6	1.1	2338	2	US-08-425-069-1
6	34.6	1.1	2338	4	US-08-317-844B-1
7	34	1.1	4368	1	US-08-026-138E-17
8	34	1.1	4446	1	US-08-026-138E-6
9	33.4	1.1	4937	1	US-08-038-682-3
10	33.4	1.1	4937	1	US-08-302-832-3
11	33.4	1.1	4937	3	US-08-530-198-3
12	33.4	1.1	4937	3	US-08-469-880-3
13	33.4	1.1	4937	4	US-08-728-470-3
14	33.4	1.1	4937	4	US-08-617-697-3
15	33.4	1.1	9323	1	US-08-038-682-6
16	33.4	1.1	9323	1	US-08-302-832-6
17	33.4	1.1	9323	3	US-08-530-198-6
18	33.4	1.1	9323	3	US-08-469-880-6
19	33.4	1.1	9323	4	US-08-728-470-6
20	33.4	1.1	9323	4	US-08-617-697-6
21	33	1.1	1753	7	5225348-2
22	32.8	1.1	1302	3	US-08-529-600D-1
23	32.8	1.1	1302	4	US-08-973-275-5
24	32.6	1.1	2277	2	US-08-676-967-5
25	32.6	1.1	2277	2	US-08-676-974-5
26	32.6	1.1	2277	4	US-09-098-487-5
27	32.4	1.1	175	1	US-08-182-175A-58

28	32.4	1.1	175	2	US-08-474-633A-76	Sequence 76, Appl
29	32.4	1.1	175	6	PCT-US92-06412-58	Sequence 58, Appl
30	32.2	1.1	1401	6	PCT-US92-06965A-32	Sequence 32, Appl
31	32	1.1	160	1	US-08-182-175A-32	Sequence 32, Appl
32	32	1.1	160	1	US-08-182-175A-34	Sequence 34, Appl
33	32	1.1	160	2	US-08-474-633A-31	Sequence 29, Appl
34	32	1.1	160	2	US-08-474-633A-31	Sequence 31, Appl
35	32	1.1	160	6	PCT-US92-06412-32	Sequence 32, Appl
36	32	1.1	160	6	PCT-US92-06412-34	Sequence 34, Appl
37	32	1.1	187	1	US-08-182-175A-90	Sequence 90, Appl
38	32	1.1	187	1	US-08-474-633A-78	Sequence 78, Appl
39	32	1.1	187	6	PCT-US92-06412-90	Sequence 90, Appl
40	31.8	1.1	660	2	US-08-307-279A-11	Sequence 11, Appl
41	31.8	1.1	2277	2	US-08-676-967-2	Sequence 2, Appl
42	31.8	1.1	2277	2	US-08-676-974-2	Sequence 2, Appl
43	31.8	1.1	2277	4	US-09-098-487-2	Sequence 2, Appl
44	31.8	1.1	2356	1	US-08-105-483-222	Sequence 222, App
45	31.8	1.1	2356	1	US-08-220-151-75	Sequence 75, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14



APPLICANT: Galcheva-Gargova, Zoya  
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX  
TITLE OF INVENTION: PROTEINS AND USES THEREOF  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/870,518  
FILING DATE: 06-JUN-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/019,219  
FILING DATE: 06-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, Peter J.  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 04020/102001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ. ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5183 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-870-518-7

Query Match 1.2%; Score 35.2; DB 4; Length 5183;  
Best Local Similarity 54.7%; Pred. No. 1.4;  
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2578 aagcaagttcaagaagttcgtcgaggttgaagtttaagacccttcaagagacaa 2637  
DB 1970 AAGACTGGTGGTGAATTCACCCAAAGTCGAAATAATTACTTTAAAGTCATGATGCC 2029  
QY 2638 cagggcctacaggtcagacgaagtaagctgaacctctcctcatccctcattgcctt 2697  
DB 2030 GAGGACTATCCGCTGATATCTCAATCTGAATCGAAGCGCATCTCTTAAATTCCTGAACCT 2089  
QY 2698 gctcatta 2705  
DB 2090 GGACTTGA 2097

RESULT 4  
US-08-371-377-16  
; Sequence 16, Application US/08371377  
; Patent No. 5851764  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Paul B.  
; TITLE OF INVENTION: DEVELOPMENT OF DNA PROBES AND  
; TITLE OF INVENTION: IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED  
; TITLE OF INVENTION: MOLECULES AND TRANSFORMATION-ASSOCIATED GENES  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America

ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/371,377  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/37590-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ. ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2128 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 621..1817  
US-08-371-377-16

Query Match 1.1%; Score 34.6; DB 3; Length 2128;  
Best Local Similarity 54.3%; Pred. No. 1.1;  
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 100 gttatagagatctggaagaaggaataacatatattcgcagggatgatggaacaa 159  
DB 1573 GTTCTGTTAAAGCTGAGAGATGGCCCTTAATCTTGAAGTCTGGATCTGCCATTTG 1632  
QY 160 ctgtatctgctgcagaggtcttggatgttaagatctgagaccagagtcacaa 219  
DB 1633 TTGATATGTTCTCTCGCAAGCCATGTGTGTGAGAGCTTCTCAGACTATCCACTTTGG 1692  
QY 220 gatcgtctg 228  
DB 1693 GCTGCTTTG 1701

RESULT 5  
US-08-425-069-1  
; Sequence 1, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25

Query Matched 1.1%; Score 34.6; DB 2; Length 2338;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0

QY 786 tctcgtgatgtagcgacaggaatgccttcctccactgaggagaagaagagtgtg 845  
| | | | |  
Db TCGTGGACAGGAGGATATGGAGGTCTTGGAAGCCAAAGTCGACAGCAGGTGATTAG 494  
| | | | |  
QY 846 atggaggagccccggagacaacattgcttgcgtctcatcaatgtgcccgaagccaatga 905  
| | | | |  
Db TGGCACAAGGTGCAGGTGCAGCACACACACCCTCGAAGTCTGTGACAAAGCCGATACGG 554  
| | | | |  
QY 906 gatatgaccttcgcaaacaccactcgtgccagagagctgtgatcacatga 954  
| | | | |  
Db TTGCTTTGGTGGACAAAGTGTCCGGACAAAGAGCATATGGAGACTTGA 603  
| | | | |

RESULT 6  
US-08-317-844B-1  
Sequence 1, Application US/08317844B  
Patent No. 59889894  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Hiltman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5989894th Washington Street  
CITY: Falls Church

Query Match	1.1%	Score 34.6	DB 4	Length 2338
Best Local Similarity	50.3%	Pred. No. 1.2		
Matches	85	Conservative	0	Mismatches 84; Indels 0; Gaps 0
Qy	786	ttctgcatgatatagacagcagagatgctcccttcctcactcagagagagaagaagatgtyg	845	
Db	435	tgcctgacacagcagcagatattgagagcttggaaagccaaagctgcacagacgagtgcatagc	494	
Qy	846	atgaggaagctccacagacacacatgctctgctgctgtctcaagaatgagcccgagccacatga	905	
Db	495	tggacaaagtgtagagtgtagacagcagcagcagccgagagtgcttgacaaagcggatagcgg	554	
Qy	906	ggatgatcccttcgcaaacaccactggcagagagactgtgatatgaactaga	954	
Db	555	tggcttggtagacaaaggtgcccggacaaagcagcattgatgagacacttgga	603	

RESULT 7  
 US-08-026-138E-17  
 ; Sequence 17, Application US/08026138E  
 ; Patent No. 5502166  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Masayoshi MISHINA  
 ; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
 ; NUMBER OF SEQUENCES: 19

```

Query Match 1.1%; Score 34; DB 1; Length 4368;
Best Local Similarity 54.9%; Pred. No. 2.9;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1072 gtttaaccagaaggagaagatcctcctcgtgacattatctcaaaagatcagtgatgt 1131
      ||| ||||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2046 GTTCACCAAGAGGGGTGTAGATGATGACCTTGCTGCTCCCTGAAAACAGGAACTTGATGC 2105

Qy 1132 agtcaattacggcaagggtgtgtctgtcacctcgtatgaaggagatgacttcggcaagt 1191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2106 ATTCACTTACGATGCAGCGCTCTCAACTACATACGTGCTGGAAAGACGAAGGCTGCAGCT 2165

Qy 1192 ag 1193
      |
Db 2166 GG 2167

RESULT 8
US-08-026-138E-6
; Sequence 6, Application US/08026138E
; Patent No. 550216
; GENERAL INFORMATION:
; APPLICANT: Masayoshi MISHINA
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
; NUMBER OF SEQUENCES: 19

```

```

Query Match      1.1%; Score 34; DB 1; Length 4446;
Best Local Similarity 54.9%; Pred. No. 2.9;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1072 gtttaaccaggagggaaggaaggaacctccctgagacatatctcaaaagatcagigaagt 1131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2124 GTTCAACCAAAAGGGGTGTAGATGATGCCCTTGCTTCCTCCGAAACCAAGGAAACTTGATGC 2183

QY 1132 agtcaattacagcaaggtgtgtgtctgcacacctgcataagaagatgacttcgcgaagt 1191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2184 ATTCAATCTACGATGAGACGCCGTGCTCACTACTAGCGCTGGAACAGACGAAGGCTGCACGT 2243

QY 1192 ag 1193
      |
Db 2244 GG 2245

RESULT 9
US-08-038-682-3
; Sequence 3, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;

```

```

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-08-038-682-3

```

```

Query Match          1.1%; Score 33.4; DB 1; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

Qy 2239 caaagattagcataccagatgacaattatgattgggaactcttcaactgctgcatga 2298
    ||||| ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 CAAAGGCTTAAACAACAGATAGAGCTCTGCAGCGGTGAATTTTAAACGCGTAATG 2468
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2299 gctggagctaatctacacacattggaagcagacataatttaaaagaccagaaactt 2358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 CAACATGTCATTCATCTCAAGAAGAGCGAAGTTAATTTCATAATTAACCAACGA 2528
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2359 gattgtctcctgaggagatttaagaattcaagtttggctc 2401
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2529 GAACATGAACACAGCAAAACCTTACCAATTCGGTTTTTAAGCC 2571
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-08-302-832-3

```

```

Query Match          1.1%; Score 33.4; DB 1; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

Qy 2239 caaagattagcataccagatgacaattatgattgggaactcttcaactgctgcatga 2298
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 CAAAGGCTTAAACAACAGATAGAGCTCTGCAGCGGTGAATTTTAAACGCGTAATG 2468
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2299 gctggagctaatctacacacattggaagcagacataatttaaaagaccagaaactt 2358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 CAACATGTCATTCATCTCAAGAAGAGCGAAGTTAATTTCATAATTAACCAACGA 2528
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2359 gattgtctcctgaggagatttaagaattcaagtttggctc 2401
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2529 GAACATGAACACAGCAAAACCTTACCAATTCGGTTTTTAAGCC 2571
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-08-530-198-3
; Sequence 3, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

```



```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-3

Query Match 1.1%; Score 33.4; DB 3; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 2239 caaagattagcatcacagatgacattatgattggaaactcttcaactcgtgcatga 2298
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2409 CAAGCGCTTAACAACACAGGTATAGAGCTCTGCAGGGGTGAATTTTACGCGCTTAATGG 2468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2239 gctgagctaatctatcacacatttgaaggcataatttaaaagacacagcaaacct 2358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2469 CAACATGTCATTCATTCCTCAAGAGGAGCGAAGTTAATTTCAAATTAACCAACGA 2528
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2359 gattgtctcctgaggagatttaagtaattcagtttgggtc 2401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2529 GAACATGAACACAGCAACACTTACCAATTCGGTTTTTAGCC 2571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-469-880-3
; Sequence 3, Application US/08469880
; Patent No. 587673
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 587673-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg

```

```

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-3

Query Match 1.1%; Score 33.4; DB 3; Length 4937;
Best Local Similarity 50.3%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 2239 caaagattagcatcacagatgacattatgattggaaactcttcaactcgtgcatga 2298
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2409 CAAGCGCTTAACAACACAGGTATAGAGCTCTGCAGGGGTGAATTTTACGCGCTTAATGG 2468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2239 gctgagctaatctatcacacatttgaaggcataatttaaaagacacagcaaacct 2358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2469 CAACATGTCATTCATTCCTCAAGAGGAGCGAAGTTAATTTCAAATTAACCAACGA 2528
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2359 gattgtctcctgaggagatttaagtaattcagtttgggtc 2401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2529 GAACATGAACACAGCAACACTTACCAATTCGGTTTTTAGCC 2571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-728-470-3
; Sequence 3, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:

```



Db 2909 CAACATGTCATTCAATCTCAAAAGAGGAGCGAAGTTAATTTCAAATTAAACCAACGA 2968  
Oy 2359 ggatttctcctgaggagatttaagtgaattcagtttgggc 2401  
Db 2969 GAACATGACACAAAGCAACCTTTACCAATTCGGTTTAAACC 3011

Search completed: April 22, 2000, 16:17:28  
Job time: 26614 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2000, 16:07:45 ; Search time 625.03 Seconds  
(without alignments)  
1206.068 Million cell updates/sec

Title: US-09-422-999-17  
Perfect score: 3013  
Sequence: 1 gatccagcgaagatgtgat.....ccacatttcaaaatgcc 3013

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues  
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.6	1.7	335	1 V90407	EST clone DL605. N
2	43.6	1.4	4062	1 T13347	C3G protein gene.
3	36.4	1.2	110000	1 V21209_12	Continuation (13 o
4	35.4	1.2	2388	1 V84554	Human secreted pro
5	35.4	1.2	2652	1 O51233	Sequence encoding
6	35.4	1.2	9601	1 N80860	Sequence corresp.
7	35.2	1.2	5183	1 V04063	Schizosaccharomyce
8	34.6	1.1	2128	1 T37405	Prostate tumour in
9	34.6	1.1	2338	1 Q14183	N.clavipes draglin
10	34.6	1.1	2338	1 V23249	Nephila clavipes s
11	34.6	1.1	4833	1 V65139	Ethiopia sp. W20
12	34.6	1.1	22243	1 V74475	Staphylococcus aur
13	34.4	1.1	1677	1 V06059	Human imidazoline
14	34.4	1.1	1954	1 X28368	Human imidazoline
15	34.4	1.1	3317	1 V06060	Human imidazoline
16	34.4	1.1	3318	1 X28369	Human imidazoline
17	34.4	1.1	3389	1 X28367	Human imidazoline
18	34.4	1.1	15202	1 X28371	Human imidazoline
19	34.2	1.1	110000	1 X20248_04	Continuation (5 of
20	34	1.1	4368	1 O55979	NMDA receptor chan
21	34	1.1	3395	1 Q56916	Glutamic acid rece
22	33.6	1.1	3395	1 V40547	Homo sapiens secre
23	33.4	1.1	1872	1 Q39220	Carrot reverse tra
24	33.4	1.1	4937	1 Q49507	Sequence encoding
25	33.4	1.1	4937	1 Q72294	Haemophilus high m
26	33.4	1.1	4937	1 T90995	Non-typeable Haemo
27	33.4	1.1	9323	1 Q49509	Gene cluster for h
28	33.4	1.1	9323	1 T90997	Non-typeable Haemo
29	33.4	1.1	9502	1 Q74770	Hepatitis C virus
30	33.2	1.1	524	1 T35109	Down-regulated sen
31	33.2	1.1	1250	1 X29775	L.lactis HsdS gene
32	33	1.1	581	1 T19018	Human gene signatu
33	33	1.1	788	1 V69012	DNA molecule encod
34	33	1.1	825	1 Q27006	HK4. Hepatitis C v

35	33	1.1	1753	1 Q06413	Sequence encoding
36	33	1.1	9360	1 Q02830	CDNA to HIV-2 RNA.
37	32.8	1.1	1302	1 T46127	Maga gene. maga ge
38	32.8	1.1	1302	1 T90222	Magnetospirillum s
39	32.8	1.1	1312	1 V27047	Xenopus neurogenin
40	32.6	1.1	1397	1 T84935	Human prostate pro
41	32.6	1.1	1797	1 T84931	Human prostate pro
42	32.6	1.1	2162	1 X14106	H. pylori GHPO 16
43	32.6	1.1	2277	1 V05372	Human telomerase p
44	32.6	1.1	2277	1 V13836	Homo sapiens mamma
45	32.4	1.1	175	1 Q37280	Clone 5-1 used in

ALIGNMENTS

RESULT 1	
ID V90407	Standard; cDNA; 335 BP.
AC V90407:	
DT 15-FEB-1999	(first entry)
DE EST clone DL605.	
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;	
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;	
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;	
KW gene therapy; ss.	
OS Homo sapiens.	
PN W09845436-A2.	
PD 15-OCT-1998.	
PE 10-APR-1998; 006955.	
PR 10-APR-1997; US-838821.	
PA (GENEX ) GENETICS INST INC.	
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,	
PI Racie LA, Spaulding V, Treacy W;	
DR WPI; 99-070077/06.	
PT New polynucleotides encoding human secreted proteins - derived from	
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,	
PT ovary, pituitary, retina and colon cDNA libraries.	
PS Claim 1; Page 530-531; 618pp; English.	
CC The present sequence represents a human expressed sequence tag (EST).	
CC The polynucleotide, which is a secreted EST, and the encoded protein	
CC are predicted to have useful biological activities which would make	
CC them suitable for treating, preventing or ameliorating medical	
CC conditions in humans and animals, although no supporting data is	
CC given. Suggested activities include nutritional activity, immune	
CC stimulating or suppressing activity, haematopoiesis regulating	
CC activity, tissue growth activity, activin/inhibin activity,	
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic	
CC activity, receptor/ligand activity, anti-inflammatory activity,	
CC cadherin/cumour invasion suppressor activity, tumour inhibition,	
CC activity. The polynucleotide may also be useful for gene therapy.	
SC Sequence 335 BP; 118 A; 53 C; 78 G; 86 T;	
SEQ	
Query Match	1.7%; Score 50.6; DB 1; Length 335;
Best Local Similarity	55.3%; Pred. No. 2e-05;
Matches 120; Conservative	0; Mismatches 94; Indels 3; Gaps 1;
QY 582 tgaatagatagaataaccacccaagaacatacagaacatgctgtggaactgac 641	
DB 79 tgaatgaagacagaaaagaaaactgactgcttccctaaatgctttttggaagcgaat 138	
QY 642 tggtagactgtagatcgaagacacacatgcttaccctccggaactaagctgttgga 701	
DB 139 ttgtcttcattgctgtt---ggaaattggaagagattcacagccctgaggaagcctgacac 195	
QY 702 tgtgcaagctcgttgaagaatggtgttcccaaccagctggaacagacacattcc 761	
DB 196 tgggacagcattatttgaataaagaaatcattcacacagttactgattgaataattca 255	
QY 762 aagaacttatttattcattcgaattcgtgattgata 798	
DB 256 AACGAGAACAGATGTTATATAGATTTCCTATGATGA 292	

WP	Sequence split into 17 fragments	LOCUS V21209	Accession	V21209
WP	Fragment Name	Begin	End	
WP	V21209_00	1	110000	
WP	V21209_01	100001	210000	
WP	V21209_02	200001	310000	
WP	V21209_03	300001	410000	
WP	V21209_04	400001	510000	
WP	V21209_05	500001	610000	
WP	V21209_06	600001	710000	
WP	V21209_07	700001	810000	
WP	V21209_08	800001	910000	
WP	V21209_09	900001	1010000	
WP	V21209_10	1000001	1110000	
WP	V21209_11	1100001	1210000	
WP	V21209_12	1200001	1310000	
WP	V21209_13	1300001	1410000	
WP	V21209_14	1400001	1510000	
WP	V21209_15	1500001	1610000	
WP	V21209_16	1600001	1664976	

Query Match	1.2%	Score 36.4;	DB 1;	Length 110000;
Best Local Similarity	49.0%;	Pred. No. 14;		
Matches 97;	Conservative 0;	Mismatches 101;	Indels 0;	Gaps 0;
QY 1316	agacttaagaacatgacccaatatctctggtcgcggagaaagctccacaggaacaga	1375		
Db 69595	AAACATATCGAAACATCATCATGTATTTATCTTAGATATATTTAACAACGGAATTA	69536		
QY 1376	gcttctaatacaagaanaactcacagcctcagcaaaagtatactgtatgcaggaacact	1435		

Query Match	Similarity	1.28;	Score 36.4;	DB 1,	Length 110000;
Best Local	Similarity 49.08;		Pred. No. 14;		
Matches 97;	Conservative	0;	Mismatches 101;	Indels	Gaps 0
QY 1316	agacttaagaacatgagccaagatgcttggtcgtgagaagctccacgacgaggaacaga				1375
DB 69595	AATCCTATCGAATAACACACATCGATGTAAATTATCTTGATATTTAACAACAGGAATTA				69536
QY 1376	gctcttaacaaagaanaactcacagcctcagcaaaagtatactgtgatacgcgaacact				1435
DB 69535	AATTAACATTAATCCCAAGACGACAGTGTTGTAATATGCAGATATTAGAGATTAAGACTTAGAT				69476
QY 1436	gaaanaaatcttgagacattctctagaacaacaaacgccttgaggcaactttaatgaagca				1495
DB 69475	GAAAAAATTAATTTTAAACATGTTGAAGTTGTATACATCAACGACGCTTAAATTAACGTT				69416

RESULT	4
ID	V84554/c
AC	V84554 standard; DNA: 2388 BP.
DT	01-MAR-1999 (first entry)
DE	Human secreted protein gene 144 clone HGLM46.
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	Homo sapiens.
PN	W09854963-A2.
PD	10-DEC-1998.
PF	04-JUN-1998; 011422.
PR	18-DEC-1997; US-070923.
PR	06-JUN-1997; US-048877.
PR	06-JUN-1997; US-048881.
PR	06-JUN-1997; US-048884.
PR	06-JUN-1997; US-048893.
PR	06-JUN-1997; US-048896.
PR	06-JUN-1997; US-048899.
PR	06-JUN-1997; US-048915.

PR 06-JUN-1997; US-048964  
PR 06-JUN-1997; US-048972  
PR 06-JUN-1997; US-049020  
PR 06-JUN-1997; US-049375



DR P-PSDB; R43578, R50990, R50991.  
 PT New peptide(s) modulating GDP exchange in complexes - with ras  
 PT protein and derived antibodies, nucleic acid etc. esp. for  
 PT diagnosis and treatment of cancer.  
 PS Claim 5; Page 16-20; 46pp; French.  
 CC The peptide(s) encoded by the sequence antagonise the interaction of  
 CC GDP-exchange factor (GRF) with the p21-GDP complex and thereby  
 CC regulate the activity of ras gene products. They can also be used  
 CC to identify other compounds which can modulate GDP exchange. The  
 CC coding sequence can be used to produce antisense products which can  
 CC inhibit oncogene expression and for use in diagnosis e.g. detecting  
 CC (over)expression of GRF for typing of cancers.  
 SQ Sequence 2652 BP; 706 A; 757 C; 629 G; 560 T;

Query Match 1.2%; Score 35.4; DB 1; Length 2652;  
 Best Local Similarity 42.5%; Pred. No. 2.8;  
 Matches 247; Conservative 0; Mismatches 331; Indels 3; Gaps 1;

QY 2220 ctttgaactatgagctccaaagattagatcccgatgacattatgatgggaac 2279  
 DB 1721 CCTTGAAGAACCTCAATGATGAGAGATCGGAGACAGCTGACCTGATACCTCG 1780  
 QY 2280 tcttcaactgctgcatgagctgagctaatctatcacacattgggaaggcataattta 2339  
 DB 1781 TCTTCAAGAAATTCCTTATGAGAGATCTTTCGACACAGATGATGAACTGGAAGA 1840  
 QY 2340 aaaaagaccacagcaacttgattgtctcctgagagatttaaatcaattcagttggg 2399  
 DB 1841 ATGAAGAGACCCCTTATGATGAGAAACCACTAAGACCTTCATGACATGTAACCTGA 1900  
 QY 2400 tctgtaactgagatctgcttctgtctcagctcagcaagcgtgtcagctatataaaat 2459  
 DB 1901 TTGCTTCAAGAAATGATGATGAGAGATGAGACATGAGGAGGATGAGGATGAGAGAT 1960  
 QY 2460 ttatttaagactgagcagccacgttaagagataaaatctgaattcttcttctcctcg 2519  
 DB 1961 GGGTGGCCGTACCTGATGATGAGAGATGAGGATGAGGATGAGGATGAGGATGAGGAT 2020  
 QY 2520 tcaatgagactaagtaacatctgctgagcagcttggcaactgagagagaactgcaca 2579  
 DB 2021 CCTGTCATGAGAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 2080  
 QY 2580 gcaagttcaagaagttcatgctgagagtttgaagtttaagacccttcaagagacaca 2639  
 DB 2081 AGCAGACTAAAGCTTGTATGATGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 2140  
 QY 2640 gggcctacagctgagcagtagcttaagctgaacctctctcatcccttcatgcttgc 2699  
 DB 2141 AGAATCTCAAGAGAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 2200  
 QY 2700 tcaatgaagatagcattactctcatgagggggaacaagacttca---ttgacaacttag 2756  
 DB 2201 ACCCTACGAGAGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2260  
 QY 2757 taaacttggaaaaaatgagcagttgatgcaaatagcagcaga 2797  
 DB 2261 TCAACTTCTCCAAATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2301

RESULT 6  
 ID N80860  
 AC N80860; standard; cDNA; 9601 BP.  
 DT 15-NOV-1990 (first entry)  
 DE Sequence corresp. to the genome of SIV isolate STLV-III Mac 142-83  
 DE in clone SIV-1  
 KW STLV-1.1; STLV-1.2; STLV-III Mac 142-83; ss.  
 OS Simian immunodeficiency virus.  
 FH key Location/Qualifiers  
 cds 551..2068  
 FT /tag= a  
 FT /gene="gag"

FT cds 1726..4893  
 FT /tag= b  
 FT /gene="pol"  
 FT 4826..5467  
 FT cds  
 FT /tag= c  
 FT /gene="g"  
 FT 5298..5633  
 FT /tag= d  
 FT /gene="x"  
 FT 5637..5939  
 FT /tag= e  
 FT /gene="r"  
 FT 5788..6084  
 FT /tag= f  
 FT /gene="tail"  
 FT 6014..6130  
 FT cds  
 FT /tag= g  
 FT /gene="art1"  
 FT 6090..8297  
 FT /tag= h  
 FT /gene="env"  
 FT 8298..8732  
 FT /tag= i  
 FT /gene="env ctd."  
 FT 8296..8391  
 FT cds  
 FT /tag= j  
 FT /gene="tat2"  
 FT 8294..8548  
 FT cds  
 FT /tag= k  
 FT /gene="art2"  
 FT 8569..9354  
 FT cds  
 FT /tag= l  
 FT /gene="f"  
 FT /gene="f"  
 FT W0805440-A.  
 FT 28-JUL-1988.  
 FT 15-JAN-1988; F00025.  
 FT 15-APR-1987; FR-005398,  
 FT (INSP) Inst Pasteur(Aliz/).  
 FT Alizon M, Montagnier L, Guetard D, Clavif F, Sonigo P, Guyader M,  
 FT Tiollais P, Chakrabarti L, Desrosiers R;  
 DR WPI; 88-220290/31.  
 DR P-PSDB; P80802, P81783, P81784, P81785, P81691, P81756, P81757, P81758,  
 DR P81759, P81760, P81761, P81762.  
 PT New peptide(s) with immunological properties of HIV-2 envelope protein -  
 PT having the structure of simian immune deficiency virus proteins,  
 PT useful in diagnosis and of vaccine components  
 PS Claim 23; Fig 1B; 86pp; French.  
 CC n80860 is contained in lambda SIV-1 which was deposited on 15/4/87 at  
 CC the CNM under numbers I-658 (PSIV-1.1) and I-659 (PSIV-1.2). It is the  
 CC cDNA to SIV-1 genomic RNA. Recombinant DNA contg. all or part of the cDNA  
 CC from this sequence inserted into a vector are claimed. Labelled fragments  
 CC of cDNA can be used as hybridisation probes. New peptides which have  
 CC immunological properties in common with those of the peptide structure of  
 CC the envelope protein of HIV-2 and also have a peptide structure in common  
 CC with that of SIV-1 glycoprotein are claimed. Antigenic and immunogenic  
 CC conjugates contg. the peptides and a kit to detect HIV-2 in biological  
 CC fluids are new. The peptides are useful for in vitro diagnosis of HIV-2  
 CC infection and some of them can be used as components of immunogens and  
 CC vaccines against HIV. Antibodies raised against them can be used for  
 CC treatment of AIDS.  
 SQ Sequence 9601 BP; 3277 A; 1809 C; 2378 G; 2136 T; 1 Others;

Query Match 1.2%; Score 35.4; DB 1; Length 9601;  
 Best Local Similarity 59.4%; Pred. No. 6.3; Indels 0; Gaps 0;  
 Matches 60; Conservative 0; Mismatches 41;

QY 1796 ttgagagaatttcaagcaaatctcagaagagatgacacacacaaagaagcagaag 1855  
 DB 8035 TTGAGGCAATATATACGGCCCTCTGATGAGAGGACCAATTTCAACAGAGAAACATG 8094  
 QY 1856 gtctttgacaagttcaatacagcgatgagagagccca 1896



DB 8095 TATGAAATTACAAAAGTGAATAGCTGGAGTGTGGCAA 8135

RESULT 7

ID V04063 standard; cDNA; 5183 BP.

AC V04063;

DE 08-JUN-1998 (first entry)

KW Schizosaccharomyces pombe RNA-binding protein ZPR1 cDNA.

KW ZPR1: RNA-binding protein; non-activated receptor complex;

KW signal transduction; epidermal growth factor receptor;

KW platelet derived growth factor receptor; cancer; tumour; marker;

KW angiogenesis; diabetic retinopathy; rheumatoid arthritis;

KW psoriasis; coronary atheroma; renal failure; gene therapy; ss.

OS Schizosaccharomyces pombe.

FH Key

FT CDS location/Qualifiers

FT 1076..2456

FT /\*tag= a

PN MO9746684-A1.

PD 11-DEC-1997.

PF 06-JUN-1997; U09911.

PR 06-JUN-1996; US-019219.

PA (UYMA-) UNIV MASSACHUSETTS.

PI Davis RJ, Galcheva-Gargova Z;

DR WPI; 98-042192/04.

DR P-PSDB; W38456.

PT Nucleic acid sequence encoding RNA-binding polypeptide ZPR1 - useful

PT to reduce angiogenesis, increase cell proliferation or kidney cell

PT regeneration or inhibit tumour growth

PS Claim 9; Fig 6; 88bp; English.

CC This nucleotide sequence comprises a cDNA clone coding for

CC Schizosaccharomyces pombe ZPR1 (see W38456), a novel protein that

CC belongs to a novel class of signalling molecules (see also

CC W38454-57) that bind to non-activated receptors (e.g. EGF and PDGR

CC receptors) and specifically bind small nucleolar RNAs (e.g. U3).

CC Saccharomyces cerevisiae, S. pombe and human ZPR1 clones (see

CC V04062-64) were identified using a mouse ZPR1 clone (see V04061) as

CC probe. The isolated clones can be expressed in the usual host

CC cells, and is a marker for growth and differentiation processes,

CC including malignant transformation of cells. ZPR1, optionally

CC expressed by gene therapy methods, can be used to inhibit the

CC proliferation of ZPR1 associated malignant cells, or for the

CC detection of ZPR1 suppressors and modulators of ZPR1 gene

CC expression. Suppressors increase cell proliferation, reduce

CC angiogenesis, e.g. to treat tumours, diabetic retinopathy,

CC rheumatoid arthritis, psoriasis and coronary atheroma, and/or

CC increase tubular regeneration of kidney cells, e.g. following acute

CC renal failure.

SQ Sequence 5183 BP; 1629 A; 906 C; 993 G; 1655 T;

Query Match 1.2%; Score 35.2; DB 1; Length 5183;

Best Local Similarity 54.7%; Pred. No. 4.9; Indels 0; Gaps 0;

Matches 70; Conservative 0; Mismatches 58;

QY 2578 aagcaagtlcaagaagtcgtatcgaggttgaagtttaagacccttcaaggaacca 2637

DB 1970 AAGACTGGTGTGAATTCACCCCAAGTCGAAATAATTAAGTCAATGATGCC 2029

QY 2638 cagggcctacagctgacagtagctaagctggaacctctcatccctcatgctt 2697

DB 2030 GAGGACTTATCCCGTATATTCGAATCTGAACCCGATCTTAAATVCTGTAACTT 2089

QY 2698 gctcatta 2705

DB 2090 GGACTTGA 2097

RESULT 8

ID T37405 standard; cDNA; 2128 BP.

AC T37405;

DT 26-NOV-1996 (first entry)

DE Prostate tumour inducing gene PRT-1.

KW Prostate tumour inducing gene-1; PRT-1; oncogene;

KW elongation factor-1 alpha; diagnosis; therapy; metastasis;

KW cell surface expression; antigen; cancer; ss.

OS Homo sapiens.

FH Key

FT primer\_bind location/Qualifiers

FT complement (317..336)

FT /\*tag= a

FT /\*tag= "L primer"

FT 567..586

FT /\*tag= b

FT /\*note= "A primer"

FT 621..1817

FT cds

FT /\*tag= c

FT /\*transl\_except= (744..746, aa:Gly)

FT /\*note= "ACC encodes Thr"

PN MO9621671-A1.

PD 18-JUL-1996.

PF 11-JAN-1996; U00307.

PR 11-JAN-1995; US-371377.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Shen R;

DR WPI; 96-342235/34.

DR P-PSDB; W03518.

PT DNA encoding prostate carcinoma tumour antigen and prostate tumour

PT inducing genes - and related proteins, vectors, antibodies, etc.,

PT for diagnosis and treatment of metastatic cancer

PS Disclosure; Fig 8A; 169pp; English.

CC Prostate tumour inducing gene-1 (PRT-1) (T37405) is a novel putative

CC oncogene that may contribute to carcinoma development in human

CC prostate and other tissues. PRT-1 was initially identified in

CC human prostatic carcinoma LNCaP cell DNA-transfected tumour-derived

CC murine CRE-Trans 6 cells using an RNA differential display method,

CC and full-length cDNA can be directly cloned from an LNCaP cDNA

CC library. The PRT-1 gene comprises a unique 5' 630 bp region and

CC a 3' truncated and mutated elongation factor-1 alpha gene. It

CC codes for a 46 kDa protein (W03518). PRT-2 (T37412), PRT-3 (T37413)

CC and a prostate carcinoma tumour antigen gene (T37414) were also

CC identified. The PRT-1 gene can be used to design probes useful

CC e.g. for detection of metastatic cancer, or to produce recombinant

CC proteins. Antisense sequences can be used in cancer therapy.

SQ Sequence 2128 BP; 608 A; 434 C; 536 G; 550 T;

Query Match 1.1%; Score 34.6; DB 1; Length 2128;

Best Local Similarity 54.3%; Pred. No. 4.3;

Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 100 gttattatgagaatctggaagaaggaataacattatctgcagaggtatattggaacaa 159

DB 1573 GTTCTGGTAAAGAGCTGGAAGATGGCCCTTAATTTCTGAAGTGTGATCTGCCTTGG 1632

QY 160 ctggtatgctgcctgcagaggtccttgatgattgaagtatcagacaggaagcagcag 219

DB 1633 TTGATATGCTTCTCGCAAGCCCATGCTGTGTGAGAGCTTCTCAGACTATTCACCTTTGG 1692

QY 220 gatgctgctg 228

DB 1693 GCTGCTTGG 1701

RESULT 9

ID Q14183 standard; cDNA; 2338 BP.

AC Q14183;

DE 15-JAN-1992 (first entry)

KW N-clavipes dragline silk protein coding sequence.

KW protein super fibre; major ampullate silk; orb web spider; ss.

OS Nephilina clavipes.

FH Key

FT cds location/Qualifiers

FT 1..2157

FT /\*tag= a



AD	V74475	standard; DNA; 22243 BP.
DT	16-MAR-1999	(first entry)
DE	Staphylococcus aureus contlg S80 ID #164.	
KW	Computer readable medium; vaccine; S.aureus infection; immunodetection; cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;	
KW	skin infection; surgical wound infection; sealed skin syndrome;	
KW	toxic shock syndrome; ds	
OS	Staphylococcus aureus.	
EH	Key	
FT	Location/Qualifiers	
FT	61. .120	
FT	/tag- a	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	1861. .1920	
FT	/tag- b	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	3661. .3720	
FT	/tag- c	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	5461. .5520	
FT	/tag- d	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	7261. .7320	
FT	/tag- e	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	9061. .9120	
FT	/tag- f	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	10861. .10920	
FT	/tag- g	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	12661. .12720	
FT	/tag- h	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	14461. .14520	
FT	/tag- i	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	16261. .16320	
FT	/tag- j	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	18061. .18120	
FT	/tag- k	
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

Query	1	gatccagcgaagatgtgatataatcttcactgcagcgaagaagttaaagctttaga	60
Db	242	GCTTCATCAACAAGGTTTAAATGATGATTCATAGATCAACAATATATTAACATTCATTT	301
Qy	61	aatttaccaccaatctccctcatcagattgctttagtggttatatagagaatcttgaaa	120
Db	302	AATGACAAAAGACTAATCGATCCGAGAGTGTGTAATTCGTACTTCAGGAGAACAAAAGAT	361
Qy	121	agggaataacatattctgcgcagggatagatg	153
Db	362	AAGTAATTTCTTGATTTGGCAAGTTTCGATATG	394

Query Match 1.1%; Score 34.6; DB 1; Length 22243; Best Local Similarity 51.6%; Pred. No. 18; Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0

```

OS Homo sapiens.
FH Key Location/Qualifiers
FT mat_peptide 1..1677
FT
FT /tag= a
FT /transl_except= (pos: 159..161, aa:Trp)
FT /transl_except= (pos: 594..596, aa:Trp)
FT /transl_except= (pos: 610..612, aa:Trp)
FT /transl_except= (pos: 646..648, aa:Trp)
FT /transl_except= (pos: 1003..1005, aa:Trp)
FT /transl_except= (pos: 1090..1092, aa:Trp)
FT /transl_except= (pos: 1147..1149, aa:Trp)
FT /transl_except= (pos: 1177..1179, aa:Trp)
FT /transl_except= (pos: 1372..1374, aa:Trp)
PN MO9731945-A1.
PD 04-SEP-1997.
PF 28-FEB-1997; U03156.
PR 20-MAY-1996; US-650766.
PR 01-MAR-1996; US-012600.
PA (UMIS ) UNIV MISSISSIPPI MEDICAL CENT.
PI Ivanov TR, Piletz JE;
DR WPI: 97-448631/41.
DR P-PSDB: W43396.
PT Nucleic acid encoding imidazoline receptor subtype 1 - useful for screening for (ant)agonist, homologous receptors, etc.
PS Claim 10; Pages 49-50; 72pp; English.
CC This sequence represents cDNA encoding a human imidazoline type 1 receptor (IR1), which has 559 (full length) or 298 (breakdown product) amino acid residues. The protein is highly unique in its sequence and can represent the first in a novel family of receptor proteins. The cDNA can be used to produce recombinant IR1, to screen for ligands binding to IR1, to raise antibodies against IR1 or to probe for other genes encoding homologous IR1 proteins.
SQ Sequence 1677 BP; 321 A; 543 C; 506 G; 307 T;

```

```

Query Match 1.1%; Score 34.4; DB 1; Length 1677;
Best Local Similarity 48.0%; Pred. No. 4.3;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```

```

QY 707 caagtcctgttagaagatggttctcaaccacgtgacagagaccattccaagac 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 CAGGCGGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726
QY 767 tttaattatctatcgattctgatatgacagagatgaccttgcctactgag 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 TTGGAATGGGGCCCCCAGACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 827 gaggaagaagaagatggtatgagagctccagagacacatgctgctgtgtcagagatg 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 767 GAGGAGGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846
QY 887 ggcgccgagccacatgagatg 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 847 GCGCGGAGAGAGACTTCTGCTG 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
X28368
ID X28368 standard; cDNA; 1954 BP.
AC X28368;
DT 18-JUN-1999 (first entry)
DE Human imidazoline receptor coding sequence clone.
KW Imidazoline receptor; human; ds.
OS Homo sapiens.
PN MO9911668-A1.
PD 11-MAR-1999.
PF 03-SEP-1997; U15695.
PR 03-SEP-1997; WO-015695.
PA (UMIS ) UNIV MISSISSIPPI MEDICAL CENT.
PI Ivanov TR, Piletz JE;
DR WPI: 99-214698/18.
DR P-PSDB: Y05243.
PT New polynucleotides encoding imidazoline receptive polypeptides -

```

```

PT useful for screening for ligands of imidazoline receptors, and for
PT isolating and identifying polynucleotides encoding imidazoline
PT receptive polypeptides
PS Claim 2; Page 59-60; 102pp; English.
CC This sequence encodes an imidazoline receptor of the invention.
CC Host cells expressing the polypeptides are useful for screening for
CC ligands of an imidazoline receptor. Restriction fragments of the
CC polynucleotide are useful as labeled probes for isolating and identifying
CC DNA material encoding polypeptides that are receptive to imidazoline
CC compounds. The isolation of polynucleotides encoding the imidazoline
CC receptive polypeptides have immunological and ligand binding properties,
CC which enable identification of agents having greater potency and/or more
CC selectivity for these receptors.
SQ Sequence 1954 BP; 367 A; 645 C; 591 G; 351 T;

```

```

Query Match 1.1%; Score 34.4; DB 1; Length 1954;
Best Local Similarity 48.0%; Pred. No. 4.7;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```

```

QY 707 caagtcctgttagaagatggttctcaaccacgtgacagagaccattccaagac 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 CAGGCGGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726
QY 767 tttaattatctatcgattctgatatgagacagagatgaccttgcctactgag 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 TTGGAATGGGGCCCCCAGACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 827 gaggaagaagaagatggtatgagagctccagagacacatgctgctgtgtcagagatg 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 787 GAGGAGGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846
QY 887 ggcgccgagccacatgagatg 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 847 GCGCGGAGAGAGACTTCTGCTG 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

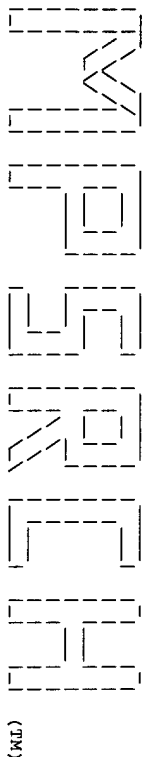
```

RESULT 15
V06060
ID V06060 standard; cDNA; 3317 BP.
AC V06060;
DT 01-MAY-1998 (first entry)
DE Human imidazoline receptor subtype 1 cDNA.
KW Imidazoline receptor subtype 1; IRL; screening; antagonist; agonist;
KM antibody; C-terminal fragment; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1365..3044
FT
FT /tag= a
FT /product= imidazoline receptor subtype 1
FT /transl_except= (pos: 1524..1526, aa:Trp)
FT /transl_except= (pos: 1959..1961, aa:Trp)
FT /transl_except= (pos: 1975..1977, aa:Trp)
FT /transl_except= (pos: 2011..2013, aa:Trp)
FT /transl_except= (pos: 2368..2370, aa:Trp)
FT /transl_except= (pos: 2455..2457, aa:Trp)
FT /transl_except= (pos: 2512..2513, aa:Trp)
FT /transl_except= (pos: 2542..2543, aa:Trp)
FT /transl_except= (pos: 2737..2739, aa:Trp)
PN MO9731945-A1.
PD 04-SEP-1997.
PF 28-FEB-1997; U03156.
PR 20-MAY-1996; US-650766.
PR 01-MAR-1996; US-012600.
PA (UMIS ) UNIV MISSISSIPPI MEDICAL CENT.
PI Ivanov TR, Piletz JE;
DR WPI: 97-448631/41.
DR P-PSDB: W43396.
PT Nucleic acid encoding imidazoline receptor subtype 1 - useful for screening for (ant)agonist, homologous receptors, etc.
PS Claim 11; Pages 50-52; 72pp; English.
CC This sequence represents cDNA encoding a human imidazoline type 1 receptor (IR1), which has 559 (full length) or 298 (breakdown product)

```







Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_dp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 22 13:20:48 2000; Maspar time 123.15 Seconds  
Tabular output not generated.  
477.985 Million cell updates/sec

Title: >US-09-422-999-18  
Description: (1-849) from US09422999.pep  
Perfect Score: 6222  
Sequence: 1 MVLKRALTMGTGFLTRTHL.....VIDNRLTSQMSHRLPRRP 849

Scoring table:  
PAM 150  
Gap 11

Searched: 225878 segs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp.archae 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mmc 8:sp.oranelle  
9:sp.phage 10:sp.plant 11:sp.todent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 53.509; Variance 96.380; scale 0.555

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	6038	97.0	1011	4	095636	CAMP-REGULATED GUANINE	0.00e+00
2	5835	93.8	993	11	0921P0	CAMP-DEPENDENT RAP1 GU	0.00e+00
3	3052	49.1	436	11	0921C7	CAMP-REGULATED GUANINE	0.00e+00
4	2942	47.3	881	4	095398	RAP1 GUANINE-NUCLEOTID	0.00e+00
5	2936	47.2	881	4	095634	CAMP-REGULATED GUANINE	0.00e+00
6	2820	45.3	884	11	0921C8	CAMP-REGULATED GUANINE	0.00e+00
7	1772	28.5	580	4	092565	MYELOBLAST KIAA0277.	0.00e+00
8	595	9.6	1499	4	09Y4G8	KIAA0313 PROTEIN.	5.32e-97
9	555	8.9	1305	5	021218	T14G10.2 PROTEIN.	3.20e-88
10	383	6.2	605	5	023867	AIMLESS RASGEF.	2.36e-51
11	370	5.9	1077	4	013905	'C3G PROTEIN'.	1.20e-48
12	351	5.6	579	4	043386	KIAA0011B (FRAGMENT).	1.02e-44
13	326	5.2	1237	4	014827	RAS-GRF2 (FRAGMENT).	1.32e-39
14	320	5.1	1189	11	070392	RAS PROTEIN-SPECIFIC G	2.17e-38
15	302	4.9	557	4	015059	KIAA0351	9.10e-35
16	302	4.9	1383	5	077086	GUANINE NUCLEOTIDE EXC	9.10e-35
17	287	4.6	489	4	016027	RAS-SPECIFIC GUANINE N	8.82e-32
18	287	4.6	1275	4	013972	RAS-SPECIFIC NUCLEOTID	8.82e-32
19	240	3.9	1048	3	012037	HYPOTHETICAL PROTEIN L	1.17e-22
20	226	3.6	347	3	014272	CAMP DEPENDENT PROTEIN	5.02e-20

21	204	3.3	689	4	094931	KIAA0846 PROTEIN.	5.52e-16
22	197	3.2	133	6	077795	CAMP-DEPENDENT PROTEIN	9.98e-15
23	192	3.1	411	2	083127	CYCIC NUCLEOTIDE BIND	7.73e-14
24	192	3.1	538	3	006729	MAT LOCUS GENES BUD5.	7.73e-14
25	194	3.1	2052	11	0921T6	FIVE FINGER-CONTAINING	3.42e-14
26	184	3.0	581	5	017869	SIMILAR TO CAMP-DEPEND	1.96e-12
27	185	3.0	593	4	015300	RET TYROSINE KINASE/CA	1.31e-12
28	180	2.9	609	4	000538	F25B3.3 KINASE LIKE PR	9.71e-12
29	172	2.8	338	5	09XTM6	CAMP-DEPENDENT PROTEIN	2.28e-10
30	175	2.8	368	5	025114	REGULATOR SUBUNIT OF	7.02e-11
31	175	2.8	434	2	P74736	CAMP PROTEIN KINASE RE	7.02e-11
32	177	2.8	638	5	019770	RAS ACTIVATOR RASGRP.	3.19e-11
33	176	2.8	795	11	0921S3	F25B3.3 PROTEIN.	4.74e-11
34	176	2.8	795	11	088469	RAS GUNYL. RELEASED P	4.74e-11
35	176	2.8	797	4	095267	CALCIUM AND DAG-REGULA	4.74e-11
36	167	2.7	325	5	094725	44 KDA REGULATORY SUBU	1.59e-09
37	165	2.7	867	5	044235	HRSH2.	3.43e-09
38	162	2.6	224	2	083286	CATABOLITE GENE ACTIVA	1.08e-08
39	154	2.5	678	5	096777	CYCIC NUCLEOTIDE AND C	2.19e-07
40	156	2.5	945	5	09Y1J9	PUTATIVE VOLTAGE-AND C	1.04e-07
41	147	2.4	1048	2	005884	HYPOTHETICAL 110.2 KD	2.87e-06
42	146	2.3	596	5	021758	SIMILARITY TO REGIONS	4.13e-06
43	143	2.3	619	2	067874	HYPOTHETICAL 71.2 KD P	1.22e-05
44	144	2.3	1080	5	023089	ZK742.1 PROTEIN.	8.50e-06
45	145	2.3	1327	4	060859	NEUROPATHY TARGET ESTE	5.93e-06

ALIGNMENTS

RESULT 1  
ID 095636: PRELIMINARY; PRT; 1011 AA.

AC 095636: 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR II.  
GN CAMP-GEFII.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eukaryota; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 99074384.  
RA KAWASAKI H., SPRINGER G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,  
RA MATSUDA M., HOUSMAN D.E., GRAIBEL A.M.,  
RT "A family of CAMP-binding proteins that directly activate Rap1.";  
RL Science 262:2275-2279(1998).  
DR EMBL: U78516; ADO3422.1; -.  
DR HSPD: P00515; 2BPB.  
SQ SEQUENCE 1011 AA; 115535 MM; 7B4462A9 CRC32;

Query Match 97.0%; Score 6038; DB 4; Length 1011;  
Best Local Similarity 99.4%; Pred. No. 0.00e+00;  
Matches 827; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db	180	TPLEHNVPLRANITTKPSEKILRAGKILRNALISRAPPHIRDRKYHKTYYRCVGT	239
Qy	18	THLIEHNVPLRANITTKPSEKILRAGKILRNALISRAPPHIRDRKYHKTYYRCVGT	77
Db	240	EIVDMNIDETPCVHSTQAVGMVYLEDGVLNHYDDEHNFODKYLFRFLDDEHDAPL	299
Qy	78	EIVDMNIDETPCVHSTQAVGMVYLEDGVLNHYDDEHNFODKYLFRFLDDEHDAPL	137
Db	300	PTEEEKCEDELDOTMLLSQMGPDAMRMILRRPGQRTVDLEIYEELHITKALSH	359
Qy	138	PTEEEKCEDELDOTMLLSQMGPDAMRMILRRPGQRTVDLEIYEELHITKALSH	197
Db	360	LSITVARELAGVILFESHAQKGTVLFENGEGSTWYIIKGSVNVYIGKGVCTLHGD	419
Qy	198	LSITVARELAGVILFESHAQKGTVLFENGEGSTWYIIKGSVNVYIGKGVCTLHGD	257
Db	420	DGKTLAVNDAAASIVIRENCHFLRYDKEDFNRIILADVANTVRLKEHDQDVLVEK	479

```

Qy 258 DEGKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVEANTVRLKEHDQDVLLEK 317
Db 480 VAGNANASOGNSOPQOKTVMGTPREKLTLEHLETTIRLEATLNATQSVLNDFTIMHCY 539
Qy 318 VAGNANASOGNSOPQOKTVMGTPREKLTLEHLETTIRLEATLNATQSVLNDFTIMHCY 377
Db 540 FMPNTQLCPALVAHYHAPOSGTGEQEKMDYALNNKRVIRLVQMAAMGDLQEDDVSM 599
Qy 378 FMPNTQLCPALVAHYHAPOSGTGEQEKMDYALNNKRVIRLVQMAAMGDLQEDDVSM 437
Db 600 ALEEFYVSDDAKRIALAEQLEPELEKIVQISEDAKAPQKKHVLLOQFNTGDERAQ 659
Qy 438 ALEEFYVSDDAKRIALAEQLEPELEKIVQISEDAKAPQKKHVLLOQFNTGDERAQ 497
Db 660 KROPINGSDEVLFKYCMQHTYTTIRVPATSVKEVISAVDKLGSGGLTIYKSSGGE 719
Qy 498 KROPINGSDEVLFKYCMQHTYTTIRVPATSVKEVISAVDKLGSGGLTIYKSSGGE 557
Db 720 KYVLKPNDSVFTTLTINGRLFACPREQFDSLPLPEQEGPTVGTGFEILMSSKDLAYQ 779
Qy 558 KYVLKPNDSVFTTLTINGRLFACPREQFDSLPLPEQEGPTVGTGFEILMSSKDLAYQ 617
Db 780 MIIYMELEFNCVHELELIYHTFGKNEFKTTANLDFLRRENIQFVVVTEICLSQSLK 839
Qy 618 MIIYMELEFNCVHELELIYHTFGKNEFKTTANLDFLRRENIQFVVVTEICLSQSLK 677
Db 840 RQQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIVSRALTWELPKFKFAEESL 899
Qy 678 RQQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIVSRALTWELPKFKFAEESL 737
Db 900 MDSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKTFEIDNLVNEFKRMJANTAR 959
Qy 738 MDSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKTFEIDNLVNEFKRMJANTAR 797
Db 960 TVRYRSOPFNDAQAANKNODVSVYROLVINDNQRTLSOMSHLEPRRP 1011
Qy 798 TVRYRSOPFNDAQAANKNODVSVYROLVINDNQRTLSOMSHLEPRRP 849

RESULT 2
ID 0921P0 PRELIMINARY; PRT; 993 AA.
AC 0921P0;
DT 01-MAY-1999 (TREMblrel. 10, Created)
RT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
OS CAMP-DEPENDENT RAP1 GUANINE-NUCLEOTIDE EXCHANGE FACTOR.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA GAUDRIAULT G.E., TAKAYA K., VALE W.W.;
RT "A brain CAMP-dependent Rap1 guanine-nucleotide exchange factor.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR HSSP; P00515; 2BPk.
SQ SEQUENCE 993 AA; 113488 MW; 52059F9A CRC32;

Query Match 93.8%; Score 5835; DB 11; Length 993;
Best Local Similarity 93.5%; Pred. No. 0.00e+00;
Matches 795; Conservative 28; Mismatches 26; Indels 1; Gaps 1;

```

```

Qy 120 DFYLFYRLDEHEDAPLPTREEKKEDEELQDTMLLSQMGPDANRMILKRPGQRTV 179
Db 324 DDELIYDELIIHKAISLSTTVAKRELAVGLIESHAKGQTVLFNOGEGTSWYIILKGS 393
Qy 180 DDELIYEEELIIKALSHLSTTVAKRELAVGLIESHAKGQTVLFNOGEGTSWYIILKGS 239
Db 384 VNVYIYKGVYCTLHEDDDGKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVE 443
Qy 240 VNVYIYKGVYCTLHEDDDGKALVNDAPRAASIVLREDNCHFLRVYKEDFNRLRDVE 299
Db 444 ANTVRLKEHDQDVLLEKVPAGNANASOGNSOPQOKTVMGTPREKLTLEHLETTIRLEPS 503
Qy 300 ANTVRLKEHDQDVLLEKVPAGNANASOGNSOPQOKTVMGTPREKLTLEHLETTIRLEPS 359
Db 504 LNEATDSVLDVFWAMCVFMPNTQLCPALVAHYHAPOSGTGEQEKMDYALNNKRVIRLV 563
Qy 360 LNEATDSVLDVFWAMCVFMPNTQLCPALVAHYHAPOSGTGEQEKMDYALNNKRVIRLV 419
Db 564 LQMAAMGDLQEDDVMAFLPEEFYVSDDAKRIALAEQLEPELEKIVQISEDAKAPQ 623
Qy 420 LQMAAMGDLQEDDVMAFLPEEFYVSDDAKRIALAEQLEPELEKIVQISEDAKAPQ 479
Db 624 KKHVLLQOENFGDERAQKROPTRGSDEVLFKYCIDHTYTTIRVPAASVKEVISAVD 683
Qy 480 KKHVLLQOENFGDERAQKROPTRGSDEVLFKYCIDHTYTTIRVPAASVKEVISAVD 539
Db 684 KLSGSEGLIYKKNSSGGEKVVLSNDVSVTTLTINGRLFACPREQFDSLPLPEQEGPT 743
Qy 540 KLSGSEGLIYKKNSSGGEKVVLPNDVSVFTTLTINGRLFACPREQFDSLPLPEQEGPT 599
Db 744 TGTVGFELMSSKDLAYQMTYDMELFNCVHELELIYHTFGKNEFKTTANLDFLRREN 803
Qy 600 TGTVGFELMSSKDLAYQMTYDMELFNCVHELELIYHTFGKNEFKTTANLDFLRREN 659
Db 804 EIOFWVTEVCLQSOLSKRVQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIVSRALTW 863
Qy 660 EIOFWVTEVCLQSOLSKRVQLKKFKIKIAHCKEYKKNLSFPAIYMGSLNIVSRALTW 719
Db 864 WEKLPSEKFKFYAEFSLMDPSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKTFE 923
Qy 720 WEKLPSEKFKFYAEFSLMDPSRNHRAVRLTVAKLEPRLPEMPLIKDMFTHEGKTFE 779
Db 924 IDNLVNEFKRMJANTARVRYRSOPFNDAQAANKNODVSVYROLVINDNQRTLSO 983
Qy 780 IDNLVNEFKRMJANTARVRYRSOPFNDAQAANKNODVSVYROLVINDNQRTLSO 839
Db 984 MSHRLPEPRRP 993
Qy 840 MSHRLPEPRRP 849

RESULT 3
ID 0921C7 PRELIMINARY; PRT; 436 AA.
AC 0921C7;
DT 01-MAY-1999 (TREMblrel. 10, Created)
RT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE 01-MAY-1999 (TREMblrel. 10, Last annotation update)
OS CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR II (FRAGMENT).
GN CAMP-GEFII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE; 99074384.
RA KAWASAKI H., SPRINGETT G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,
RT MATSUDA M., HOUSMAN D.E., GRAYBIEL A.M.;
RT "A family of CAMP-binding proteins that directly activate Rap1.";
RL Science 282:2275-2279 (1998).
DR EMBL; U78517; AAD03423.1; -.
FT NON_TER 1
SQ SEQUENCE 436 AA; 50117 MW; 30F35FB2 CRC32;

```



Query Match 49.1%; Score 3052; DB 11; Length 436;  
 Best Local Similarity 95.6%; Pred. No. 0.00e+00;  
 Matches 417; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

DB 1 RVIRLVLYQMAAMGDLQEDVDVAMAFLEERYVSSDDARMMAVAFKQALAEKTVQKQISE 60  
 QY 414 RVIRLVLYQMAAMGDLQEDVDVAMAFLEERYVSSDDARMMAVAFKQALAEKTVQKQISE 473  
 DB 61 DAKAPQKKHVVLLQOQFNTGDERAKROPINGSDVLEFKVYCIDHTDTITVPAVASVKEV 120  
 QY 474 DAKAPQKKHVVLLQOQFNTGDERAKROPINGSDVLEFKVYCIDHTDTITVPAVASVKEV 533  
 DB 121 ISAVADKLGSSEGLIIVKMSGGEKVLLKPNDSVFETTLINGRLPACPREQDSLPPLP 180  
 QY 534 ISAVADKLGSSEGLIIVKMSGGEKVLLKPNDSVFETTLINGRLPACPREQDSLPPLP 593  
 DB 181 EQBEPPTGVGTGTEFLMSKDLAVQMTYDWEELFNCVLELELIYHTGRRHFKKTTANL 240  
 QY 594 EQBEPPTGVGTGTEFLMSKDLAVQMTYDWEELFNCVLELELIYHTGRRHFKKTTANL 653  
 DB 241 FLRFENFIQFVYVTEICLSQSLSKRVQLLKCKIAHCKEYKNLSFPGIVAGLSVAV 300  
 QY 654 FLRFENFIQFVYVTEICLSQSLSKRVQLLKCKIAHCKEYKNLSFPAIVAGLSNIAV 713  
 DB 301 SRLATWEKLPKSKFKFYAFESIMPSRNHKAAYRLTAKEPLLPMPMLLKDMTFTH 360  
 QY 714 SRLATWEKLPKSKFKFYAFESIMPSRNHKAAYRLTAKEPLLPMPMLLKDMTFTH 773  
 DB 361 EGKNTFIDNLVNEKEMIMANTARTVYRYSOPFNDPAQANKNHDVRSYQOLNVIND 420  
 QY 774 EGKNTFIDNLVNEKEMIMANTARTVYRYSOPFNDPAQANKNHDVRSYQOLNVIND 833  
 DB 421 QRTLSQMSHRLPRRP 436  
 QY 834 QRTLSQMSHRLPRRP 849

RESULT 4  
 ID 095398 PRELIMINARY; PRT; 881 AA.  
 AC 095398;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
 DE RAP1 GUANINE-NUCLEOTIDE EXCHANGE FACTOR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MUSCLE;  
 RX MEDLINE; 99068645.  
 RA DE ROOIJ J., ZWARTKRUIS F.J.T., VERHEIJEN M.H.G., COOL R.H.,  
 RA NUYMAN S.M.B., WITTINGHOFF A., BOS J.L.;  
 RT "Rap1 is a Ras1 guanine-nucleotide-exchange factor directly activated  
 RT by cyclic AMP.";  
 RL Nature 396:474-477(1998).  
 DR EMBL; A0103905; AAC83381.1; -.  
 DR HSSP; P00515; 2BPX.  
 SQ SEQUENCE 881 AA; 99312 MW; 43A29191 CRC32;

Query Match 47.3%; Score 2942; DB 4; Length 881;  
 Best Local Similarity 51.1%; Pred. No. 0.00e+00;  
 Matches 406; Conservative 189; Mismatches 188; Indels 11; Gaps 8;  
 DB 30 RWPPLINSESLDFSELEQASTERYLRAGROLRHRLATCPNLTRDKRYLRYQCCS 89  
 QY 16 RRTHLIPHPVLRPANTITKVPSEKILRAGKILRNALISAPRMINDRKYHLTKYQCCV 75  
 DB 90 GRELVDDIALGLGYSRSOVVGLCOVLDEGALCHYKHKHMAQODRACQYRRPGEPP 148  
 QY 76 GTELVDMMDIETPCVHSRTQAVGMMOVLLEDVNLNHDVDEHHPDF-YLTYRFLDEHED 134

DB 149 -EVGTHE--ME--ETIAEAVALLSQRGPDALLVALRKPCGRTDEDLIFEELHKA 203  
 QY 135 APLPFEKKEDEDELQDTMLLSQMGPDANHRMLLRKPCGRTYDDLEIITEELHKA 194  
 DB 204 VAHLSNVKRELAAILFEPSKAGTVLFSQDKSTWYIIWKGSVNVYTHGKGLVTLH 263  
 QY 195 LSHLSTVARELAGVILFESHAKGTVLFEQGEESTWYIIILKGSVNVYTHGKGLVTLH 254  
 DB 264 EGDDEGOLALVNDAPRAATIIIREDNCHFLRYDKODFNRITIKDVAKMRLREHKKVLY 323  
 QY 255 EGDDEGOLALVNDAPRAATIIIREDNCHFLRYDKODFNRITIKDVAKMRLREHKKVLY 314  
 DB 324 LERASQGA-GPSPPTPGNRRTVMSGTEPEKLEIFELPAMGPDSSAHPTFETLSDFLT 382  
 QY 315 LKVPAGNNAASQNSQPOQKTVMSGTEPEKLEIFELPAMGPDSSAHPTFETLSDFLT 374  
 DB 383 HAVFMSAQLCAALLHHFHPVBPAGSGEDRSTYVCKRQOILRLVSQWVALYGSMLHTDP 442  
 QY 375 HCVFMPNTQLCPALVAHYHAPSGQTEQEKMDYALNNKRRVIRLVLYQMAAMGDLQEDD 434  
 DB 443 VATSFLQKSLDVLGHDTRLNLLREQWPERRCHLENGCCNASQMAKRNLPWLPNOD 502  
 QY 435 VMAFLFEERYVSSDDARMIAALKEDLPLEKIVQISE-DAKAPQKKHVVLLQOQFNTG 493  
 DB 503 EPLPGSSCAIOGVGDVYPYDICRDSVLTLOLPYASREVMAALAOEDGWTKGQVLYV 562  
 QY 494 ERAQQRQ-PIRSDEVLKRVYCMHTYTTIRYPVATSKVEYISAVADKLGSSEGLIIVK 552  
 DB 563 NSAGDAIGLQDPARQVASTGLNERLFFVNPQEAHELIPHPDOLGPTVSAEGDLIVSAK 622  
 QY 553 SSGGEKVLLKPNDSVFETTLINGRLPACPREQDSLPPLPQEBPVTGVGTGTEFLMSK 612  
 DB 623 DIAGOLTHDMSLFNSIHQVELIHYVLPQHLRDVTTANLEFRMRRELQYVATELCL 682  
 QY 613 DLAYQMTYDWEELFNCVLELELIYHTGRRHFKK-TTANLDELFRFENFIQFVYVTEICL 671  
 DB 683 CPVPGPRAQLKFKFKIAAHLKEOKNLSFPAVMSGNSAISRAHMERLPKRVRLY 742  
 QY 672 CSQSLSKRVQLLKCKIAHCKEYKNLSFPAIVAGLSNIASTRALATWEKLPKSKFKFY 731  
 DB 743 SALERLDPSWNRHYRLALAKISPPVLPFMPMLLKDMTFTHGNTLVENLNEKMRM 802  
 QY 732 AAFESIMPSRNHKAAYRLTAKEPLLPMPMLLKDMTFTHGNTLVENLNEKMRM 791  
 DB 803 MARAARMHHCSSH 816  
 QY 792 IANTARTVYRYSQ 805

RESULT 5  
 ID 095634 PRELIMINARY; PRT; 881 AA.  
 AC 095634;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
 DE CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR I.  
 GN CAMP-GEFI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 99074384.  
 RA KAWASAKI H., SPRINGETT G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,  
 RA MATSUDA M., HOUSMAN D.E., GRAYBIEL A.M.;  
 RT "A family of CAMP-binding proteins that directly activate Rap1.";  
 RL Science 282:2275-2279(1998).  
 DR EMBL; U78168; A012740.1; -.  
 DR HSSP; P00515; 2BPX.  
 SQ SEQUENCE 881 AA; 99376 MW; 8C1C3C82 CRC32;

Query Match 47.2%; Score 2936; DB 4; Length 881;  
 Best Local Similarity 50.9%; Pred. No. 0.00e+00;

Matches 404; Conservative 190; Mismatches 189; Indels 11; Gaps 8;		
Db	30	RMPRLNSESLEDFSESLQASTERTLRAGROHLHLLATCNLLRDKRYHLRQCCS 89
Qy	16	RRHLLEPHVPLRPANTITKVPSEKILRAGKILRNALISRAPHMIRDRKYHLKTYRCCV 75
Db	90	GRELVDGILALGIVSHSQAVGICQVLLDEGALCHVKHDMAFODRDAQFYRFPPEPE- 148
Qy	76	GTELVDWMIDEPCHVSRQAVGMQVLLDEGVLNHVDEHNFQDF-YLFYRFLDEHND 134
Db	149	-PVGTNE--ME--EELAEVALLSQGRDALLTVALKRPPGQRTDELDLIFEELHKA 203
Qy	135	APLPEEKKCEDELDOTMLLSOMGRDAMRMILKRPQRTVDLELITYEELHKA 194
Db	204	VAHLSNRYKRELAVALLEPHSKAGTVLFSQDGKTSWYIIKGSVNVYTHKGVLVTLH 263
Qy	195	LSHLSTYKRLAGVLIFFESHAGKTVLFRNQGEGTSWYIIKGSVNVYTHKGVCSTLH 254
Db	264	EGDDEFGQALVNDAPRAATITLREYNCHFLRVDKODFNRIKRDVEAKTYRLEHGKVLY 323
Qy	255	EGDDEFGKIALVNDAPRAASIVLRDNCHFLRVDEKEDFNILRDVEANTVRLKEHDQVLY 314
Db	324	LEPASOGA-GPSRPTPGRRNRYTVMSGTPDKILELLLEAMGLDSSAHNDKFTPLSLIT 382
Qy	315	LEKVPAGNRSNAGNSQPOOKYTVMSGTPEKILEHLEFLTRLEATLNEATDVLNDETIM 374
Db	383	HREFMASQCALHHLHFIVERPAGSEORSTYVCNKRQOILRLYSQWALGSMILHDP 442
Qy	375	HCFVMPNQLCSRALVANYHNAQPSQGTQEKMDALNNKRRVRLVQWAMAMGDLQEDD 434
Db	443	VATSFHLKSLDLYGRDTRLSNLRQWPERRCHRLNCGNASPQKARNLPVMLPNOD 502
Qy	435	VSMATLEEFYVVSDDARMIALKEQLRELEKIVKOISE-DAKARQKKHVVLLQDFNGD 493
Db	503	EPLPGSCAIIQVGDKVPYDICRPHSVTLQDLPYASVREYMAALAOEDGWTGQVLYK 562
Qy	494	ERAKQRO-PINGSEVLEFYVCMDHTYTTIRVPATSVKEVISAADKLGSGEGILLIM 552
Db	563	NSAGDAIGLOPDAVGATSLGNRIEFLVVPNOEHLIRPHROLGPYVSGAGLDLVSK 622
Qy	553	SSGGEVVLKPNVSVFTTLTINGRLRACRQDFDSLPRQDEGTVGTGTFELMSK 612
Db	623	DLAQGLTDHWSLFSIHQVELIHVGLRQHLRDVTTANLEFRMRFENLQYVATELCL 682
Qy	613	DLAQMTIYDWELFNCVHELELIYHFGHNFKK-TTANLDLFRFRENIOGVVYTELCL 671
Db	683	CPVPGRAQLLKFTKLAHLKEQKVNVSFEAVMGLSNPSISLAHTWERLPHKVRKLY 742
Qy	672	CSQLSKRVOLKKFKITIAHCKEYKNLSNFFAIVMGLSNAVSRLATLWEKLPSEKFEKY 731
Db	743	SALERLDPSMNRHYRLALAKTSPVPIRPMLLKDMFTIEGNTLVENLINFEKMM 802
Qy	732	AEEESLMDPSRNHRAYRLTVAKLEPRLPMLIKDMFTIEGNTFIDNLVNEFKMM 791
Db	803	MARARMLHCRSH 816
Qy	792	IANTARTVRYRSQ 805
RESULT 6 PRELIMINARY; PRT; 884 AA.		
ID	Q921C8	AC
Qy	01-MAY-1999 (Tremblrel. 10, Created)	
DT	01-MAY-1999 (Tremblrel. 10, last sequence update)	
DE	01-NOV-1999 (Tremblrel. 12, last annotation update)	
OS	CAMP-REGULATED GUANINE NUCLEOTIDE EXCHANGE FACTOR I.	
GN	CAMP-GEFI.	
OC	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 99074384.	

RA	KAWASAKI H., SPRINGETT G.M., MOCHIZUKI N., TOKI S., NAKAYA M.,
RA	MATSUDA M., HOSUMI D.E., GRAVIEL A.M.;
RT	"A family of CAMP-binding proteins that directly activate Rap1.";
RL	Science 282:2275-2279(1998).
DR	EMBL: U78167; RAD12739.1; -.
DR	HSSP: P00515; 2BPX.
SO	SEQUENCE 884 AA; 100256 MW; C8BB3814 CRC32;
Query Match 45.3%; Score 2820; DB 11; Length 884;	
Best Local Similarity 50.1%; Pred. No. 0.00e+00;	
Matches 401; Conservative 186; Mismatches 200; Indels 14; Gaps 10;	
Db	30 RMPRLNSESGLDFVSESLQASTERTLRAGROHLHLLATCNLLRDKRYHLRQCCS 89
Qy	16 RRTHLLEPHVPLRPANTITKVPSEKILRAGKILRNALISRAPHMIRDRKYHLKTYRCCV 75
Db	90 GRELVLDGILALGIVSHSQAVGICQVLLDEGALCHVKHDMAFODRDAQFYRFPPE- 146
Qy	76 GTELVDMIDETPCVSHSRQAVGMQVLLDEGVLNHVDEHNFQDF-YLFYRFLDDEHND 134
Db	147 -PQVAG--THVEEELYEAMALLSQGRDALLTVALKSGQRTDELDLIFEELHKA 203
Qy	135 APLPEEKKCEDELDOTMLLSOMGRDAMRMILKRPQRTVDLELITYEELHKA 194
Db	204 VAHLSNRYKRELAVALLEPHSKAGTVLFSQDGKTSWYIIKGSVNVYTHKGVLVTLH 263
Qy	195 LSHLSTYKRLAGVLIFFESHAGKTVLFRNQGEGTSWYIIKGSVNVYTHKGVCSTLH 254
Db	264 EDDDDGQALVNDAPRAATITLRENNCHFLRVDKODFNRIKRDVEAKTYRLEHGKVLY 323
Qy	255 EDDDDGKIALVNDAPRAASIVLRDNCHFLRVDEKEDFNILRDVEANTVRLKEHDQVLY 314
Db	324 LEPTSGA-GPSRPTPGRRNRYTVMSGTPDKILELLLEAMGLDSSAHNDKFTPLSLIT 382
Qy	315 LEKVPAGNRSNAGNSQPOOKYTVMSGTPEKILEHLEFLTRLEATLNEATDVLNDETIM 374
Db	383 HSEVMPSTQLEALLHNFHVERPESAGSEDERSTYVCNKRQOILRLYSRWALYSPMLR 442
Qy	375 HCFVMPNQLCSRALVANYHNAQPSQ--GTQEKMDYALNNKRRVRLVQWAMAMGDLQ 431
Db	443 SRPVATSFLOKSLDLYSDTRLSNLRQWPERRCHRLNCGNASPQKARNLPVMLPNOD 502
Qy	432 EDDVSMATLEEFYVVSDDARMIALKEQLRELEKIVKOISE-DAKARQKKHVVLLQDFN 490
Db	503 HNEEPLPSAGAIRGDKVPYDICRPHSVTLQDLPYASVREYMAALAHEDHNTKGQVLY 562
Qy	491 TQDERAQKRP-PKRSDEVLEKRYVCMDHTYTTIRVPATSVKEVISAADKLGSGEGILLI 549
Db	563 VKVSAGDVVGLQPDARGVATSLGNRIEFLVVPNOEHLIRPHROLGPYVSGAGLDLV 622
Qy	550 VKMSGGKVVLYKPNVSVFTTLTINGRLRACRQDFDSLPRQDEGTVGTGTFELM 609
Db	623 SAKDLAQGLTEHWNLEFNRIHQVELIHVGLRQHLRDVTTANLEFRMRFENLQYVATE 682
Qy	610 SSKDLAQMTIYDWELFNCVHELELIYHFGHNFKK-TTANLDLFRFRENIOGVVYTE 668
Db	683 LCICPVPGRQQLKFTKLAHLKEQKVNLSFPAVMGLSNPSISLAHTWERLPHKVR 742
Qy	669 ICLCSQSKRVOLKKFKITIAHCKEYKNLSNFFAIVMGLSNAVSRLATLWEKLPSEK 728
Db	743 KLYSALERLDPMSNRHYRLATKLSPPVIRPMLLKDMFTIEGNTLVENLINFEK 802
Qy	729 KYFAEFESLMDPSRNHRAYRLTVAKLEPRLPMLIKDMFTIEGNTFIDNLVNEFK 788
Db	803 MRMARARMLHCRSHSTAP 823
Qy	789 MRMIANTARTAVYRSQPNP 809
RESULT 7 PRELIMINARY; PRT; 580 AA.	
ID	Q92565
AC	Q92565;
DI	01-FEB-1997 (Tremblrel. 02, Created)

SO SEQUENCE 580 AA; 67733 MW; 1CA9F77D CRC32;  
DR PRAM; PF00617; RASGEF; 1.  
DR PRAM; PF00618; RASGEF; 1.  
DE MYELOBLAST KIAA0277.  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 97191544.  
RA NAGASE T., SEIT N., ISHIKAWA K., OHIRA M., KANARABAYASI Y., OHARA O.,  
RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIA0201-KIA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
DR EMBL; D87467; BA013406.1; -  
DR PRAM; PF00617; RASGEF; 1.  
DR PRAM; PF00618; RASGEF; 1.  
DE MYELOBLAST KIAA0277.  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 97191544.  
RA NAGASE T., SEIT N., ISHIKAWA K., OHIRA M., KANARABAYASI Y., OHARA O.,  
RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIA0201-KIA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
DR EMBL; D87467; BA013406.1; -  
DR PRAM; PF00617; RASGEF; 1.  
DR PRAM; PF00618; RASGEF; 1.  
DE MYELOBLAST KIAA0277.

[illegible]

```

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A..
RC TISSUE-BRAIN.
RX MEDLINE, 97342984.
RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N., TANAKA A., KOYANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL: AB002311; BAA20772.1; -.
SQ SEQUENCE 1499 AA; 167416 MW; DTGB0593 CRC32;.

Query Match          9.6%; Score 595; DB 4; Length 1499;
Best Local Similarity 36.6%; Pred. No. 5,32e-97;
Matches 112; Conservative 78; Mismatches 101; Indels 15; Gaps 13;

Db 606 PDVLV-RVKKAQOGRSMISKDTPAKEVVYIARFAYATAPDDYSCLCEVSTPEGVIK 664
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 505 SDEVLEKRYVCMDHTTTRRVAVSAVEK--ISAVAD-KL-GSGDELIIVKKSSGGEKVY- 560
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 665 QRRLPDQLSKLADIQLDSGRYYILKNMETETLCSDDAOEELRESQILDOL-STVEYAT 723
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 561 LK--PNDVSVEFTT-LTINGRIFLACPRQFPDLAPE-OEGPTVTGGVFIEIMSSKLAY 616
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 724 QLSNRNFELFNRIEPTLEYIDDLF-KLSKTSKANLRFEVINQGFNVASILLRETNOL 782
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 617 QMTIYWELFNCVHHELLHYTFPGRHNPKTTINDLFLRRENGIQFWYVEI-CLESOL 675
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 783 -KRKKTIKEFIKIALHCCKCNFSMSFAISGLNAPVARLTWKBPKNYEKIFODLO 841
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 676 SKERVOLLKFIRKIAHCKEYKNLNSPFAIWMGLSNLAWSLALTWERKPSRKFFVEFE 735
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 842 DLDPDSRNAKRRNVJNSQNLQPPILIFPVIIKDITFLHGDSKVGLVNEFKLMRIA 901
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 736 SLMDPSRNHRVR--LVTAKLEPLPIPFMPLIKDMTFHKGKTFIDNLVNFEKMRIA 793
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 902 KEIRHV 907
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 794 NTARTV 799
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT      9      PRELIMINARY:           PRT: 1305 AA.
ID ID021218
AC AC021218; Q22503;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE DE114G10.2 PROTEIN.
GN GN114G10.2.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A..
RA WILD A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z69664; CAA93519.1; -.
DR EMBL: Z68880; CAA93519.1; JOINED.
DR EMBL: Z68880; CAA93100.1; -.
DR DR EMBL: Z69664; CAA93100.1; JOINED.
SQ SEQUENCE 1305 AA; 144639 MW; BI30E596 CRC32;.

Query Match          8.9%; Score 555; DB 5; Length 1305;
Best Local Similarity 36.3%; Pred. No. 3,20e-88;
Matches 90; Conservative 71; Mismatches 80; Indels 7; Gaps 6;
```





OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 97349984.  
 RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,  
 RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL; AB002349; BAA20808.1; -.  
 DR PFAM; PF00617; RasGEF; 1.  
 SO SEQUENCE 557 AA; 62132 MW; 4CB4D1BA CRC32;

Query Match 4.9%; Score 302; DB 4; Length 557;

Best Local Similarity 25.8%; Pred. No. 9,10e-35;  
 Matches 64; Conservative 66; Mismatches 108; Indels 10; Gaps 10;

Db 44 FDVLKVTPEEFASQITLMDIPVFKAIQPEELASCGWSKKEKHSIADPNVVAFTFRNQVSF 103  
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 Qy 606 FE-L-MSSKRLAYQMTIYDWELENCVHELELIYTFGRHNFKKTANLDLFLRFNEIQF 663  
 Db 104 WVVREI-LTAQTLKIRAEILSHFVKIAKKLLELNTHSLMSVSAIQSAPIFRLTKTAL 162  
 ||| ||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 Qy 664 WVTVEICLSQ-LSKRVQLKKFKIKIAHCKEYKLNLSFPAIYVGLSNIASVRLALTWEK 722  
 Db 163 LNRDKTFPEKIDYLMKSEKDYKRTREYIRSLKMWPSIPYLGIVLIDLIIYDSAVPASGS 222  
 | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 Qy 723 LPSKFKKTYAEFESLMDPSRNNHRAVRLTVAKLE-PLIPFMPLIKDMFTHEGKNTFID 781  
 Db 223 IMENQORSNOMNILLRIADLQVSCSYDLTTLPHVQKYLKSVR-YIEELQKVEDDNY- 280  
 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 Qy 782 NLVVFEEKRMIAATFVRYRSQ-PFNP-DA-AQANKNHODVRSYVROLNVINDNQRTLS 838  
 Db 281 KLSLRTEP 288  
 : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 Qy 839 QMSHRLRP 846

Search completed: Sat Apr 22 13:22:56 2000  
 Job time : 128 secs.



```

OY 92 SRTQAGVMQVLLDEGVLNHNVDGHHFQD-FYLFREFLDEHEDAPLFTPEEKKCEDEEL 150
DB 528 SKSITLLSSVABEPLFLMTIVSKPGEERSEPELEVVYEEILFKALSHLSTMYKROLNMFV 587
OY 151 ODTMLISMGCDANHRMLRKPGRVDDDEIIEEILHFKALSHLSTMYKRELACVL 210
DB 588 KYEQVYHAGSVYFROGEIVYVYIVKGAVEYNVNGK-IVCLRGDGDGKALVNDLPR 646
OY 211 IRESNAKGGTVLFNQEESTSWITLKGSVNVIYTKGVSNDLHGBDQGLATVNDLPR 270
DB 647 AATVYTEDDSMFLVVDKHFHNOILHOEANTVRLKDGEDVLLEKYDIPRGALLENSEN 706
OY 271 AASIVLRDNCHFLRVYDKEDFNRLRDVEANTVRLKEHQDVLVEKY--PAGNASNQG 328
DB 707 SCNFNGSVYMAKAKKILEVLETER-RIDALDDISLDEVEDVETILTHDAPMDNTVCN 765
OY 329 NSOPOOKYVMGTPREKLEHLEFETIRLEATLNEATD-SVL-NDEIMHCVFMPTOLCP 386
DB 766 FLKSYFFTPYATRDISTDSCTEEVRCKRVYQVYVWCSSLRVNFFL-NPVTNSFYEE 824
OY 387 ALVANHHAOPSGTEDEKMD-YALN-N-KRVIYRLVLOMAAY-GDLLOEDDVSMATLEE 442
DB 825 LFCHVLDKRRKLGEMDILFRIGSI-RSTRENMOVLARHPAIV--LDCGVLSAHTPCPV 881
OY 443 FYVYSDDARMIALAEQLEPELEKIVKQISEDAKAPQKKHVKLLQOFMTGDERAQRQPI 502
DB 882 LPSDVCNOIYIADTTCPFLPIRVOKTAIEICELSRKRSFSAEPLNVEYKSNKEKILF 941
OY 503 RGSDEVLEFVYCMHDYTTTIRVPVATSVKEVISAADAKIG-SGEGILLIYKSSGSEKVL 561
DB 942 SPNDRAIPVLINSLKLYVYVNEEIPFLVPMEDONGPTSSHSLIHLIDSOELAHOLEL 1001
OY 562 KRNDVSVEFTLTITNGLELFCPREQEDSLTPREQBPVYGTGT-ELMSSDLAYQMTI 620
DB 1002 PHQLLRSTDSNELLVYVIGRESFPLSMFNLDLLVRFENEYOHMSTTEILLATE-ENRM 1060
OY 621 YWELFNCVHELELYHFGFGRHNFKTTA-NDLFLRRENEIOFWVWEICLSQSLKRV 679
DB 1061 ELIKKFIATATAREYRDLITVFAITLGLSHTSISRLITMSKLPASIKTFSELENLID 1120
OY 680 QLLKFKFIKAACHCKEKNLNSFALVMGSLNIAVSRLALVTEKRLSPKRRKYAEESLMD 739
DB 1121 PPRNHRMYRLVLYSKMSPYIPFVPLIKDLMFIHOGNKSFEVGLVNFPMHFAKIFRSF 1180
OY 740 PSRNHRAYLITVAKLEPPLIPMLIKDMTTHGCKNTFTDNLVNEFKMIANTARTV 799
DB 1181 ROCKSQ-MONGAHEFIEPQSLIRMLR 1206
OY 800 RYRSQPFMPDAQAQANKNHQDVRSYVR 826

```

RESULT 2  
ID CC25\_SACKL STANDARD: PRT: 1095 AA.  
AC Q02342;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CELL DIVISION CONTROL PROTEIN 25 (FRAGMENT).  
GN CDC25.  
OS Saccharomyces kluyveri (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
CC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92354938.  
RA PRIGOSY T., GONZALES E., BROEK D.;  
RT "Identification and analysis of a DNA fragment from Saccharomyces  
kluyveri that can complement the loss of CDC25 function in  
Saccharomyces cerevisiae".  
RL Gene 117:67-72(1992).  
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS  
PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,

```

CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M82964; AAA34479.1; -.
DR PIR: PC1114; PC1114.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PRAM: PF00617; RASGEF; 1.
DR PRAM: PF00618; RASGEF; 1.
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane.
FT NON_TER 1
FT TRANSMEM 1
SQ SEQUENCE 1095 AA; 125605 MM; 765862 CRC32;

```

Query Match 5.18; Score 317; DB 1; Length 1095;  
Best Local Similarity 32.5%; Pred. No. 1,63e-37;  
Matches 67; Conservative 50; Mismatches 81; Indels 8; Gaps 6;

```

DB 808 LLDLSDIDYAKQLIRKHSFLFKISPECDRTWGNKYCMGSGSKNTETISNHLTN 867
OY 608 LMSKRLAY-QMTIYWELEFNCVHELELYHFGFGRHNFKTTA-NDLFLRRENEIOFW 664
DB 868 VSFMIYVQTDIKRIQLOPFIVNVAACHLNNFSSLTALISALYSSPIRLKRTWAAYP 927
OY 665 VVTEICLSQSLSKRVOLKFKFIKAACHCKEKNLNSFALVMGSLNIAVSRLALVTEKLP 724
DB 928 EYKRLLELNTLMDSAKNFIRYROLKLSIGDPCVPEFVYVSLTFANGNPDLHRN 967
OY 725 SKFKFYAEFESLMDPSRNHRAYR-LTVAKLEPPLIPFMLIKDMTTHGCKNTFTD-N 782
DB 968 TVLVNFGKRVILLEIKESIVYQRSH 1013
OY 783 --LVNFEKMIANTARTVRYR-RSQ 805

```

RESULT 3  
ID GMRP\_RAT STANDARD: PRT: 1244 AA.  
AC P28818;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GMRP) (P140 RAS-GRF).  
GN RASGRF1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-BRAIN:  
RX MEDLINE: 92350260.  
RA SHOU C., FARUSWORTH C.L., NEEL B.G., FEIG L.A.;  
RT "Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing  
factor for Ras p21.";  
RL Nature 358:351-354(1992).  
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



RN [3]

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

```

OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90190870.
RA HUGHES D.A., FUKUI Y., YAMAMOTO M.;
RT "Homologous activators of ras in fission and budding yeast.;"
RL Nature 344:355-357(1990).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN IS ESSENTIAL FOR MATING.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53254; CAA37345.1; -.
DR PIR: S28098; S28098.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00617; RASGEF; 1.
DR PFAM: PF00618; RASGEF; 1.
DR PFAM: PF00618; RASGEF; 1.
KW Guanine-nucleotide releasing factor; SH3 domain.
FT DOMAIN 1
SQ SEQUENCE 911 AA; 105185 MW; 8C5BC852 CRC32;

Query Match 4.5%; Score 278; DB 1; Length 911;
Best Local Similarity 22.2%; Pred. No. 9,01e-30;
Matches 57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;

DB 638 QLELDLS-LHNSPDIYKDELIVLLPPREIAKQCLIEFOSFHSIRIQFLTKIMDLN 696
QY 584 EGFSDTLPLEQEGPTVGTVGTFELMSSKDLAYQMTIYDWELFNCVHELEIYHTFGHN 643
DB 697 RSPKRTSTFELSN-HLVNF-VTEIYOEBRRRTAVLAFVLOVCYLAEINFAFLF 754
QY 644 -FKTTANLDLFRFPNETIQFWVTEICLSQLSKRVOLKFKIAHCKEYKNLSFF 702
DB 755 STISALNSSPIHRLKRTWANLNSKTASFELINLTLEARKNFSNYRDLCEVCVPL 814
QY 703 AIVMGISNAVSRLALTWKLPKSKFKFYAEFESLMDSRNHRARLVAKLEPPLIPFM 762
DB 815 GYFFDLPFLTKGNKDNFQNMINDKRTKRVIRILNEIKFQSGVGYMNPINEVOELLE 874
QY 763 PLLIKDMFTFHEGNKTFIDNLVNEFKMRMIANTARTVRYRSQPPNDAQANKNH-ODV 821
DB 875 ISRENTNNI-YORSLT 890
QY 822 KSYVROLNVIDNQRTLS 838

RESULT 6
AC S052_HUMAN STANDARD; PRT; 1332 AA.
ID 007890; 015503;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SON OF SEVENLESS PROTEIN HOMOLOG 2 (SOS-2).
GN SOS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN:
RX MEDLINE: 9326494.
RA CHARDIN P., CAMONIS J.H., GALE N.W., VAN AELST L., WIGLER M.H.,
RA BAR-SAGI D.;

```

```

RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2.;"
RL Science 260:1338-1343(1993).
RN [2]
RP SEQUENCE OF 724-1296 FROM N.A.
RX TISSUE-PLACENTA:
RA FATH I., APIOU F., DUTRILLAUX B., TOCQUE B.;
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13858; AAA35914.1; -.
DR EMBL: L20686; AAA91852.1; -.
DR MIM: 601247; -.
DR PROSITE: PS00720; GDS_CDC25; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00617; RASGEF; 1.
DR PFAM: PF00621; RASGEF; 1.
DR PFAM: PF00621; RASGEF; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 200 441
FT DOMAIN 442 546
FT DOMAIN 546 561
FT DOMAIN 561 575
FT DOMAIN 575 758
FT DOMAIN 758 1180
FT DOMAIN 1180 1208
FT DOMAIN 1203 1208
FT CONFLICT 778 778
FT CONFLICT 861 861
FT CONFLICT 948 948
FT CONFLICT 948 948
FT CONFLICT 999 999
FT CONFLICT 1032 1032
FT CONFLICT 1042 1042
FT CONFLICT 1112 1114
FT CONFLICT 1159 1159
FT CONFLICT 1296 1296
SQ SEQUENCE 1332 AA; 153030 MW; 9C70BC7B CRC32;

Query Match 4.4%; Score 272; DB 1; Length 1332;
Best Local Similarity 25.8%; Pred. No. 1,33e-28;
Matches 59; Conservative 70; Mismatches 90; Indels 10; Gaps 7;

DB 752 FESPPPIEWHISKPGQFETDMLDPIEIAQRLLESIDLKVKQPSLVGSVTKED 811
QY 586 FDSITPLPEQEGPTVGTVGTFELMS--SKDLAYQMTIYDWELFNCVHELEIYHTFGHN 643
DB 812 KEINSPILKMIHTNTLTLTFEKCIVEANFEERVAVLSRIEILIOVODLNFGVLE 871
QY 644 FKTTANLDLFRFPNETIQFWVTEICLSQLSKRVOLKFKIAHCKEYKNLSFF 703
DB 872 IVSAVNSVSYRRDHFEEALQERRKRL-D-EA-VESODHEFKYLVKLSINPCVPFF 928
QY 704 IVMGLSNAVSRLALTWKLPKSKFKFYAEFESLMDSRNHR-RAVLLTVAKLEPPLIPFM 762
DB 929 GIVLTNLTKEGNNDFLKRKGLINFSKRRVAETGEIQOYQNPY 977
QY 763 PLLIKDMFTFHEGNKTFI--D-NLVNFEKMRMIANTARTVRYRSQPP 807

RESULT 7
AC S052_MOUSE STANDARD; PRT; 1297 AA.
ID 002384;

```





QY 671 LC-SQSKRVQLKFKFIKIAHCKEYKNNLSFPAIVMGLSNIAVSRLATWKEKPSKFK 729

Db 1189 NIKNMKNKLNSSRNENEYR-DVUKFGSEPCYPPFEGVITSDITFVYHGNPDLYNKTQY 1247

QY 730 FYAEFSLMDPSRNHAFYRLTYAK-LEP-PLIPFMDLKKMTFHEGNKTFI-D-N-LV 784

Db 1248 NFAKRAKTSRIVSGIDRFKTKGTGYNFOEVEIQK 1280

QY 785 NFEKRMIAINTARTVRYRYSQPPN-PDAQAQAK 816

RESULT 11

ID CC25\_YEAST STANDARD; PRT; 1589 AA.

AC P04821;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-JAN-1988 (Rel. 06, last sequence update)

DT 15-JUL-1999 (Rel. 38, last annotation update)

DE CELL DIVISION CONTROL PROTEIN 25.

GN CDC25 OR CINI OR YLR310C OR L2142.6.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 87131091.

RA BROEK D., TODA T., MICHAELI T., LEVIN L., BIRCHMEIER C., ZOLLER M., POWERS S., WIGLER M.;

RT "The S. cerevisiae CDC25 gene product regulates the RAS/adenylylate cyclase pathway."

RL Cell 48:789-799(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 86220116.

RA CAMONIS J.H., KALEKINE M., GONDRE B., GARREAU H., BOY-MARCOITE E., JACQUET M.;

RT "Characterization, cloning and sequence analysis of the CDC25 gene which controls the cyclic AMP level of Saccharomyces cerevisiae."

RL EMO J. 5:375-380(1986).

RN [3]

RP SEQUENCE FROM N.A.

RX STRAIN-8288C / AB972;

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z., FAVELLO A., FULTON L., GATTONG S., GRECO T., KIRSTEN J., KICCAIA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D., JOHNSON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENZES S., MILLER N., NHAN M., PAULEY A., PELUSO D., RIEKEN L., RILES L., TATCH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;

RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.

RN [4]

RP DOMAINS.

RX MEDLINE: 89181526.

RA MUNDER T., MINK M., KUNTZEL H.;

RT "Domains of the Saccharomyces cerevisiae CDC25 gene controlling mitosis and meiosis."

RL Mol. Gen. Genet. 214:271-277(1988).

RN [5]

RP FUNCTION.

RX MEDLINE: 91203884.

RA JONES S., VIGNAIS M.L., BROACH J.R.;

RT "The CDC25 protein of Saccharomyces cerevisiae promotes exchange of guanine nucleotides bound to ras."

RL Mol. Cell. Biol. 11:2641-2646(1991).

CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START, THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X03579; CAA27259.1; -

DR EMBL: M15458; AAA34478.1; -

DR EMBL: U17247; AAB67360.1; -

DR EMBL: U20618; AAB64528.1; -

DR PIR: A26596; RGRYC5.

DR HSSP: P29355; ISEM.

DR SGD: L0000263; CDC25.

DR PROSITE: PS00720; GDS\_CDC25; 1.

DR PROSITE: PS50002; SH3; 1.

DR PFAM: PF00018; SH3; 1.

DR PFAM: PF00617; RASGEF; 1.

DR PFAM: PF00618; RASGEF; 1.

KW Guanine-nucleotide releasing factor; Cell division; Cell cycle; Mitosis; Transmembrane; SH3 domain.

FT TRANSMEM 1452 1473 POTENTIAL.

FT DOMAIN 58 128 SH3.

FT CONFLICT 497 497 I -> Y (IN REF. 2).

FT CONFLICT 954 963 PYGHPEPKN -> LSVYIMNLSR (IN REF. 2).

SQ SEQUENCE 1589 AA; 179091 MW; A28FA88A CRC32;

Query Match 4.18; Score 255; DB 1; Length 1589;

Best Local Similarity 25.58; Pred. No. 2.55e-25;

Matches 71; Conservative 77; Mismatches 117; Indels 13; Gaps 10;

Db 1228 VSEKFGAEDLLQKINEKINEKEPVVQKODSVSNVQTKRDNKSPI-HMS5-SSL 1285

QY 537 VADKLGSGEGLI--IV-KMSGGEKVLKPNV-SVFETTLINGSLFACPREQDSLPL 592

Db 1286 PSSASSAFPRKRLKLLDPTTYATQTLVLEHDIYLRITPECLDRMGKRYCNMGSP 1345

QY 593 PEGEGPVGTGVTELMSSKDLAY--QMTIYWELENCVHELELIYHFGHNRKT-TA 649

Db 1346 NIKETLANNTLNFESHTIVKQADYKTSKLTQYFVTVACHKELNFSMTAIVSALY 1405

QY 650 NDLFLRFNFEIOPFWVTEICLSQSKRVOLKFKFIKIAHCKEYKNNLSFPAIVGLS 709

Db 1406 SSPIYLRKKTLDIVSTESDLKNNLNMDSKRNFVKYRBLRSYTDVACYPFEGVYLS 1465

QY 710 NIAVSRLATWKEKPSKFKFYAEFSLMDPSRNHAFYRLTYAKLEPPLIPFMDLKK 768

Db 1466 LTFEPWGNPEFLNSNININFSKRTIANIVEIISFK 1503

QY 769 MTFTHEGNKTFI-D--NLVNFEMKMIANTARTVRYR 803

RESULT 12

ID SOS\_DROME STANDARD; PRT; 1595 AA.

AC P26675;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, last sequence update)

DT 15-JUL-1999 (Rel. 38, last annotation update)

DE SON OF SEVENTEEN PROTEIN.

GN SOS.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-ONEGON-R;

RC MEDLINE: 92141820.

RA BONFINI L., KARLOVICH C.A., DASGUPTA C., BANERJEE U.;

RT "The Son of seventeen gene product: a putative activator of Ras."

RL Science 255:603-606(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92034591.

RA SIMON M.A., BOWTELL D.D.L., DODSON G.S., LAVERY T.R., RUBIN G.M.;

RT "Ras1 and a putative guanine nucleotide exchange factor perform



	RESULT	14	KAPR_APICA	STANDARD:	PRT:	377 AA.
ID	AC	Dt	p1319;			
Dt	DT	01-JUL-1993	(Rel. 26, Created)			
Dt	DT	01-JUL-1993	(Rel. 26, Last sequence update)			
Dt	DE	15-DEC-1999	(Rel. 39, Last annotation update)			
Dt	CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN (N4 SUBUNIT OF PROTEIN KINASE A).					
OS	Aplysia californica (California sea hare).					
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidaea; Aplysidae; Aplysia.					
RN	[1]					
RX	SEQUENCE FROM N.A.					
RA	BEROLD P.J., BEISHAUSEN S.A., SACKTOR T.C., CHELEY S., BAYLEY H., SCHWARTZ J.H.;					
RT	"A regulatory subunit of the CAMP-dependent protein kinase down-regulated in aplysia sensory neurons during long-term sensitization."					
RL	Neuron 8:387-397(1992).					
CC	-1- SUBUNIT: THE INACTIVE FORM OF THE ENZYME IS COMPOSED OF TWO REGULATORY CHAINS AND TWO CATALYTIC CHAINS. ACTIVATION BY CAMP PRODUCES TWO ACTIVE CATALYTIC MONOMERS AND A REGULATORY DIMER THAT BINDS FOUR CAMP MOLECULES.					
CC	-1- PWM: THE PSEUDOPHOSPHORYLATION SITE BINDS TO THE STRATE-BINDING REGION OF THE CATALYTIC CHAIN BUT IS NOT PHOSPHORYLATED. THE PHYSIOLOGICAL SIGNIFICANCE OF PHOSPHORYLATIONS BY OTHER KINASES IS UNCLEAR.					
CC	-1- SIMILARITY: CONTAINS 2 CYCLIC NUCLEOTIDE-BINDING DOMAINS.					
CC	-1- SIMILARITY: BELONGS TO THE CAMP-DEPENDENT KINASE REGULATORY CHAIN FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; X62382; CAI44246.1; -.					
DR	PIR; JH0590; OKGAKI.					
DR	HSSP; P00514; IBBK.					
DR	PROSITE; PS00889; CAMP_BINDING_1; 2.					
DR	PROSITE; PS00889; CAMP_BINDING_2; 2.					
DR	PROSITE; PS50042; CAMP_BINDING_3; 2.					
DR	PFAM; PF00027; CAMP_binding; 2.					
KW	CAMP-binding; phosphorylation; Repeat; Acetylation.					
FT	INIT MET	0				
FT	DOMAIN	1	132			
FT	SITE	92	96			
FT	NP_BIND	133	250			
FT	NP_BIND	251	377			
FT	REPEAT	133	250			
FT	REPEAT	251	377			
FT	MOD_RES	1	1			
FT	MOD_RES	97	97			
FT	BINDING	198	198			
FT	BINDING	207	207			
FT	BINDING	322	322			
FT	BINDING	331	331			
FT	SEQUENCE	377 AA;	42606 MW;	5F481A94 CRC32;		
Query Match	Best Local Similarity	3.5%;	Score 216;	DB 1;	Length 377;	
Matches	33; Conservative	26;	Mismatches 39;	Indels 3;	Gaps 3;	
Dn	134 FSHLDNRSQIFDM-PVVRHAGEVIITOGDEGDNFYVDIGEVDDYY-NNHVHSISG 191 ::    : : :   :   :   :    :   :   :   : :					
Oy	195 LSHISTYVKRELAVLTIESHAKEGTIVLPNDGEESTSYTIILKSQNVNYITKGVCVLH 254					
Dn	192 EGGSFGELALTYGPRAATVKAETDV-KIWMDIRDSYRRIL 231					

Qy	255	EGDDFGKALVNDAPRAASTIVLR	EDNC	HLRVKDC	EDFNRL	295
RESULT	15					
ID	KAPR.CAEEL	STANDARD:		PRT:	376 AA.	
AC	P30625; Q21920;					
DT	01-APR-1993 (Rel. 25, Created)					
DT	15-DEC-1993 (Rel. 39, Last sequence update)					
DT	15-DEC-1999 (Rel. 39, Last annotation update)					
DE	CAMP-DEPENDENT PROTEIN KINASE REGULATORY CHAIN.					
GN	KIN-2 OR KIN-A OR R07E4.6.					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nemata; Secernentea; Rhabditia; Rhabditidae;					
OC	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-BRISTOL N2;					
EX	MEDLINE; 90153982.					
RA	LU X., GROSS R.E., BAGCHI S., RUBIN C.S.;					
RT	"Cloning, structure, and expression of the gene for a novel					
RT	regulatory subunit of cAMP-dependent protein kinase in Caenorhabditis					
RT	elegans.";					
RL	J. Biol. Chem. 265:3293-3303(1990).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-BRISTOL N2;					
RA	MILLER N.;					
RL	Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.					
CC	-1 SUBUNIT. TETRAMER, COMPOSED OF 2 REGULATORY (R) AND 2 CATALYTIC					
CC	(C) SUBUNITS. IN THE PRESENCE OF CAMP IT DISSOCIATES INTO 2 ACTIVE					
CC	MONOMERIC C SUBUNITS AND AN R DIMER.					
CC	-1 PTM: THE PSEUDOPHOSPHORYLATION SITE BINDS TO THE SUBSTRATE-					
CC	BINDING REGION OF THE CATALYTIC CHAIN BUT IS NOT PHOSPHORYLATED.					
CC	THE PHYSIOLOGICAL SIGNIFICANCE OF PHOSPHORYLATIONS BY OTHER					
CC	KINASES IS UNCLEAR.					
CC	-1 SIMILARITY: CONTAINS 2 CYCLIC NUCLEOTIDE-BINDING DOMAINS.					
CC	-1 SIMILARITY: BELONGS TO THE CAMP-DEPENDENT KINASE REGULATORY CHAIN					
CC	FAMILY.					
CC	-----					
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).					
CC	-----					
CC	EMBL; J05220; AAA27980.1; -					
DR	EMBL; U39652; AAA80405.1; -					
DR	PIR; A35076; OKMIR.					
DR	HSSP; P00514; IAPR.					
DR	WORMPEP; R07E4.6; CE04821.					
DR	PROSITE; PS00888; CNMP_BINDING_1; 2.					
DR	PROSITE; PS00889; CNMP_BINDING_2; 2.					
DR	PROSITE; PS50042; CNMP_BINDING_3; 2.					
DR	PRAM; PF00027; CNMP_binding; 2.					
DR	CAMP-binding; Phosphorylation; Repeat.					
FT	DOMAIN	1	131			
FT	SITE	92	96			
FT	NP_BIND	132	249			
FT	NP_BIND	250	376			
FT	REPEAT	132	249			
FT	REPEAT	250	376			
FT	DISULFID	23	23			
FT	DISULFID	44	44			
FT	MOD_RES	97	97			
FT	BINDING	197	197			
FT	BINDING	206	206			
FT	BINDING	321	321			
FT	BINDING	330	330			
FT	BINDING	330	330			

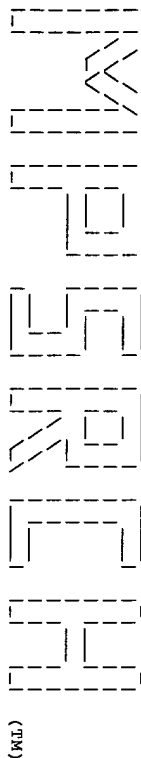
FT CONFLICT 214 214 T -> I (IN REF. 1).  
FT CONFLICT 260 260 A -> V (IN REF. 1).  
SQ SEQUENCE 376 AA; 42609 MM; B4F7833B CRC32;

Query Match 3.4%; Score 209; DB 1; Length 376;  
Best Local Similarity 29.7%; Pred. No. 9.48e-17;  
Matches 33; Conservative 31; Mismatches 44; Indels 3; Gaps 3;

Db 133 FAHLEDEQKTYDAM-FPVEKSAGETIIEGEGEDNFYIDKGTVDYVYNHEVYL-TIN 190  
Qy 195 LSHLSTYVKRELAVGLIFESHAKGTVLENQEGSTWYIILKGSVNVVIYGRGVCTLH 254  
Db 191 EGGSPGELALYGPRAATVIKTIV-KLMAIDRLTYRIILMSVTKRKRM 240  
Qy 255 EGDDEGKLALVNDAPRAASIVLREDNCHFLRVDKEDENRIILDYEANTVRL 305

Search completed: Sat Apr 22 13:20:31 2000  
Job time : 54 secs.





Release 3.1a John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 22 13:18:40 2000; MasPar time 36.15 Seconds

Tabular output not generated. 941.038 Million cell updates/sec

Title: >US-09-422-999-18  
Description: (1-849) from US09422999.pep  
Perfect Score: 6222  
Sequence: 1 MVLKRALTMGTGFLTRTHL.....VIDNQRLSQMSHRLPRRP 849

Scoring table:  
PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 53.983; Variance 117.292; scale 0.460

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description	Pred. No.
1	1922	30.9	1213	2	S42368	guanine nucleotide re 0.00e+00
2	317	5.1	1095	2	PC1114	SKDC25 protein - yea 6.73e-31
3	310	5.0	1244	2	S29083	guanine-nucleotide-re 9.95e-30
4	291	4.7	1260	2	S28407	guanine nucleotide-ex 1.40e-26
5	287	4.6	1275	2	A38985	nucleotide exchange f 6.36e-26
6	278	4.5	911	2	S28098	ste6 protein - f1sio 1.89e-24
7	266	4.3	1297	2	S25714	son-of-sevenless-2 pr 1.67e-22
8	255	4.1	1333	2	A37488	Ras guanine nucleotid 9.72e-21
9	253	4.1	1333	2	S30356	CDC25 protein homolog 2.03e-20
10	255	4.1	1336	2	S25716	Ras guanine nucleotid 9.72e-21
11	255	4.1	1589	1	RGBYC5	cell division control 4.22e-20
12	251	4.0	1596	2	A41216	guanine nucleotide ex 2.32e-18
13	240	3.9	1048	2	S64758	SCD25 protein (versio 9.82e-18
14	236	3.8	1250	2	S14177	SCD25 protein (versio 9.82e-18
15	216	3.5	378	1	OKGARI	protein kinase (EC 2. 1.21e-14
16	210	3.4	1435	1	BVBYL1	guanine nucleotide-re 9.89e-14
17	208	3.3	376	1	OKKMR	protein kinase (EC 2. 1.98e-13
18	205	3.3	377	1	OKKMR	protein kinase (EC 2. 5.60e-13
19	207	3.3	381	1	A60669	protein kinase (EC 2. 2.80e-13
20	204	3.3	381	1	OKMSR1	protein kinase (EC 2. 7.91e-13
21	206	3.3	411	2	S18634	cgsl protein - f1sio 3.96e-13
22	203	3.3	416	1	OKBYRC	protein kinase (EC 2. 1.12e-12
23	201	3.2	379	1	OKHUR1	protein kinase (EC 2. 2.22e-12

RESULT ENTRY TITLE	1	ALIGNMENTS
S42368	#type complete	probable cyclic nucle 4.76e-11
guanine nucleotide releasing factor homolog - Caenorhabditis elegans		BDU5 protein kinase - yeast 6.68e-11
07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 07-Aug-1998		2.9 378 1 OKBO1R 2.62e-09
ACCESIONS S42368		2.9 381 1 OKHUR1 1.88e-09
REFERENCE S42368		2.9 381 1 OKHUR1 2.62e-09
#authors Smith, A.		2.9 381 1 OKHUR1 2.62e-09
#submission submitted to the EMBL Data Library, March 1994		2.9 381 1 OKHUR1 2.62e-09
#accession S42368		2.9 381 1 OKHUR1 2.62e-09
#status preliminary		2.9 381 1 OKHUR1 2.62e-09
#molecule_type DNA		2.9 381 1 OKHUR1 2.62e-09
#residues 1-1213 #label SMI		2.9 381 1 OKHUR1 2.62e-09
#cross-references EMBL:230423; NID:g458479; PID:g458480		2.9 381 1 OKHUR1 2.62e-09
GENETICS #introns 196/3: 238/1; 263/3; 312/3; 444/3; 519/3; 550/2; 669/2; 742/1; 776/3; 1072/2; 1102/3; 1107/3		2.9 381 1 OKHUR1 2.62e-09
CLASSIFICATION #superfamily cAMP receptor protein cyclic nucleotide-binding domain homology; CDC25-type guanine nucleotide exchange activator homology		2.9 381 1 OKHUR1 2.62e-09
FEATURE 574-689		2.9 381 1 OKHUR1 2.62e-09
987-1206		2.9 381 1 OKHUR1 2.62e-09
SUMMARY		2.9 381 1 OKHUR1 2.62e-09
Query Match 30.9%; Score 1922; DB 2; Length 1213;		2.9 381 1 OKHUR1 2.62e-09
Best Local Similarity 38.5%; Pred. No. 0.00e+00;		2.9 381 1 OKHUR1 2.62e-09
Matches 311; Conservative 204; Mismatches 259; Indels 33; Gaps 24;		2.9 381 1 OKHUR1 2.62e-09
DB 414 EKQIRDSGILRHKMLTDHQVIRDTTEHTRVQCMGAEIMDLTLFVSTGSSSL 473		2.9 381 1 OKHUR1 2.62e-09
OY 39 EKILR-AGKILNALISRAPIHMRKYLKTYRCCVGTLEVDMM-----ID-ETPCVH- 91		2.9 381 1 OKHUR1 2.62e-09
DB 474 SRIOKSAIWOVLNGLSHIDGHOFLDKTNSYRW---OQ---FSSRNKVASIEV 527		2.9 381 1 OKHUR1 2.62e-09
OY 92 SFTQAVGMVQVLLDEGVNLNHDQENHFOF-EYLFYRFIDDEHEDAPLPTEEKKCEDDEL 150		2.9 381 1 OKHUR1 2.62e-09
DB 528 SKSTILSSVAPETFLFMVSKSGFERSPDELEVYEELTFKASHLSTWKKRLSNFV 587		2.9 381 1 OKHUR1 2.62e-09
OY 151 QDTMLLSQMGDPDAHMRMLTKRPGQRTYVDLEITTEELHLTKALSHLSTYKRLAGVL 210		2.9 381 1 OKHUR1 2.62e-09
DB 588 KYEQYVHAGSVYFROGELISVYVYIVLKGAVENVNGK-IVCILIRGDEFGKILATVNDPR 646		2.9 381 1 OKHUR1 2.62e-09

```

QY 211 IFESHAKGTVLFPNOGEESTSWYIILKGSVNVVYIGKVGVCTLHBDDEFGKLTALVNDAPR 270
DB 647 AATVYTEDDSMEFLVYDKHFNQIILHOEVANTVRLKDGEDVLVEKYEDIRGALAEASN 706
QY 271 AASIVLRDNCFLRYDKDEFNKILRDVANTVRLKEHDQDLVEKY--PAGNPNASNOG 328
DB 707 SCNFGVSVMAKAEKILEVLET-RIDALGDDISELDFEDEDILTTHAFMPDNTVCN 765
QY 329 NSQPOQKYVMSGTEPKILEHLEFLETRLEATLENATD-SVL-NDIIMHCVFMPRTQJCP 386
DB 766 FLKSYIFRPIRYATIDSTIDSTEERCKRNVQVYVWCSLLRNFFL-NDPTNSVEE 824
QY 387 ALVAHNAOPSGTDEKMD-YALN-N-KRVIYRLVQWAMV-GDLLOEDDVSAFLEE 442
DB 825 LFCVHYDDKRLGGMEDILTRIGST-RSTRVMQVLARHPATV-LDQGLSANTPCPY 881
QY 443 FYVSYSDDAKMTAALKEOLPELEKTVKQISEDAKAPQKKHVKYLQOFMTGDERAKROP 502
DB 882 LPSDVCONIIVLADTYTCFLPIRVDTAEICELSRMRMSFAPPLNVEVKSNGELIF 941
QY 503 RGSDEVLFRVYCMQDHYTTIRVPVATSVKEVISAVALKLG-SGEGILLIVKMSGGEKYYL 561
DB 942 SPNDRAIPVLSLNSKLYVNRBEIPLVPMEDQNGPTPSHSHSLIHLIDSOELAHQFL 1001
QY 562 KENDSVFTTLINGRLFCAPREQFDLSLPLPEQEGPTVGVGTF-ELMSSKDLAYQMTI 620
DB 1002 FHLQILRSDSNELLVYGRSESPLSMPFNLDLVRRENVQHSSTEILLATE-ENRM 1060
QY 621 YMELEFNCVHELELTYHFFGRHNEFKKTTA-NLDLFRNFNELOFVWVEICISQSLSRV 679
DB 1061 ELIKREFISATTAAREYRDLTFVATTLGSHTSISRLTLTWSKLPSPALKTFSLELND 1120
QY 680 QLKPFIRKIAAHCXEKKNLSFFALVMGSLNVAISRLALTMWKLPSRKRTYAESESIMD 739
DB 1121 PPRNHRMRLVSKMSPTIPFPVPLIKDMFTIHOQKNSFTYGLNFEKMHAKIFERSF 1180
QY 740 PPRNHRMRLVAKLEPPLIPMLIKDMFTIHOQKNSFTYGLNFEKMHAKIFERSF 799
DB 1181 ROCKSO-MDNGAEHEFIEPOSTLRMLR 1206
QY 800 RYRSQPFNPDAQAQANKHODVRSYR 826

RESULT 2
ENTRY PC1114 #type fragment
TITLE SKDC25 protein - yeast (Saccharomyces kluyveri) (fragment)
ORGANISM #formal_name Saccharomyces kluyveri
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
09-Sep-1997
ACCESSIONS PC1114
REFERENCE PC1114
#authors Prigozy, T.; Gonzales, E.; Broek, D.
#journal Gene (1992) 117:67-72
#title Identification and analysis of a DNA fragment from
Saccharomyces kluyveri that can complement the loss of
CDC25 function in Saccharomyces cerevisiae.
#cross-references MVID:92354938
#accession PC1114
#molecule_type DNA
#residues 1-1095 #label PRI
#cross-references GB:M82964; NID:g171186; PID:g171187
GENETICS
#gene SKDC25
CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator
#homology homology
KEYWORDS transmembrane protein
FEATURE
808-1049 #domain CDC25-type guanine nucleotide exchange activator
homology #label SOS
SUMMARY #length 1095 #checksum 2353

Query Match 5.1%; Score 317; DB 2; Length 1095;
Best local similarity 32.5%; Pred. No. 6.73e-31;

```

```

Matches 67; Conservative 50; Mismatches 81; Indels 8; Gaps 6;
DB 808 ILDDISLDYAKQLTRKHSFLFKISPECDLRWGNKCYCMGSKNTEPISNHLTVN 867
QY 608 LMSSKDLAY--QMTTYMELFNCVHELELTYHFFGRHNEFKKTTA-NLDLFRNFNIQFW 664
DB 868 VSFMIIVQOTDKRIQIOLQFPIVNAVAHCHELNPFSSITAFISASPPIRLKRTMAAVP 927
QY 665 VMEICISQSLSKRVOLKFKIRIAAHCXEKKNLSFFALVMGSLNVAISRLALTMWKL 724
DB 928 EYKKLLEINTLMDSAKNRIRYRLQKLSIGDEPCVPEGVYSLDTFTANGNPDFLHRN 987
QY 725 SKFKFYAEFESIMDPSRNHRAVR-LTVAKLEPPLIPFMLIKDMFTIHOQKNSFTY 782
DB 988 TVLVNFKRVRILLEIKESIVYQRSH 1013
QY 783 --LVNFKRMRIANTARTVRY--RSQ 805

RESULT 3
ENTRY S29083 #type complete
TITLE guanine-nucleotide-releasing protein - rat
ALTERNATE_NAMES CDC25 protein homology
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
17-Mar-1999
ACCESSIONS S29083
REFERENCE S29083
#authors Shou, C.; Farnsworth, C.L.; Neel, B.G.; Fely, L.A.
#journal Nature (1992) 358:351-354
#title Molecular cloning of cDNAs encoding a
guanine-nucleotide-releasing factor for Ras p21.
#cross-references MVID:92350260
#accession S29083
#molecule_type mRNA
#residues 1-1244 #label SHO
#cross-references EMBL:X67241; NID:957664; PID:957665
#note the authors translated the codon GAG for residues 135
and 137 as Gln
CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator
#homology homology; CDC24 homology; pleckstrin repeat homology
FEATURE
240-426 #domain CDC24 homology #label CD24\
1005-1241 #domain CDC25-type guanine nucleotide exchange activator
homology #label SOS
SUMMARY #length 1244 #molecular-weight 14266 #checksum 2921

Query Match 5.0%; Score 310; DB 2; Length 1244;
Best local similarity 26.4%; Pred. No. 9.95e-30;
Matches 65; Conservative 74; Mismatches 98; Indels 9; Gaps 7;
DB 1005 FENHSAEIAEQLTLDHLVFKSIPYEEFGQGMKADKNERPIYIMKTRHNSNLI 1064
QY 606 FELMSSDLDAYQMTTYMELFNCVHELELTYHFFGRHNEFKKTTA-NLDLFRNFNIQFW 665
DB 1065 ASEL-LRNEEVSARASTIEKVAADICRCLHYNVAVLETSSINSAIFRLKRTWLKYS 1123
QY 666 VTEICLDSQ--LSKRVOLKFKIRIAAHCXEKKNLSFFALVMGSLNVAISRLALTMWKL 724
DB 1124 KQTKSLFDKLOKLVSSDGRKRLRETLRNCDCPCVPLGLNYLTDLAFLEGTPNNTEDQL 1183
QY 725 SKFKFYAEFESIMDPSRNHRAVR-LTVAKLEPPLIPFMLIKDMFTIHOQKNSFTY 783
DB 1184 VNSKPMISHIIRERQOQTYKLEP-QP-KVYQL--VDETVLDE-SLEASLR 1237
QY 784 VNEFKMIRIANTARTVRYRSQPFNPDAQAQANKHODVRSYQLVINDORTLSQMSHR 843
DB 1238 IEPKLP 1243
QY 844 LEPRRP 849

RESULT 4

```

RESULT	5	
ENTRY	A38985	#type complete
TITLE	nucleotide exchange factor CDC25 - human	
ALTERNATE_NAMES	Ras-specific human nucleotide-releasing factor	
ORGANISM	#formal name Homo sapiens #common name man	
DATE	19-Jul-1996	#sequence_revision 19-Jul-1996 #text_change 10-Sep-1997
ACCESSIONS	A38985; A46199; I58371	
REFERENCE	A38985	
#authors	Wei, W.; Broek, D.	
#submission	submitted to GenBank, December 1994	
#description	Cloning and analysis of the full length human cdc25 cDNA, a Ras-specific nucleotide exchange factor.	
#accession	A38985	
##status	translated from GB/EMBL/DBJ	
##molecule-type	mrna	
##residues	1-1275	#label WEI
##cross-references	GB:L26564; NID:g433719; PID:g433720	
REFERENCE	A46199	
#authors	Wei, W.; Mosteller, R.D.; Sanyal, P.; Gonzales, E.; McKinney, D.; Dasgupta, C.; Li, P.; Liu, B.X.; Broek, D.	
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:7100-7104	
#title	Identification of a mammalian gene structurally and functionally related to the CDC25 gene of Saccharomyces cerevisiae.	
##cross-references	WUID:92357779	
#accession	A46199	
##molecule-type	mrna	
##residues	1047-1054, 'A', 1056-1112, 'G', 1114-1135, 'C', 1137-1275	
##experimental_source	brain	
##note	sequence extracted from NCBI backbone (NCBIN:111098, NCBI:P:111099)	
REFERENCE	I58371	
#authors	Schweighoffer, F.; Faure, M.; Fath, I.; Chevallerier-Multon, M.C.; Apion, F.; Durrillaux, B.; Sturani, E.; Jaquet, M.; Toque, B.	
#journal	Oncogene (1993) 8:1477-1485	
#title	Identification of a human guanine nucleotide-releasing factor (H-GRF5) specific for Ras proteins.	
##cross-references	WUID:93275641	
#accession	I58371	
##status	preliminary; translated from GB/EMBL/DBJ	
##molecule-type	mrna	
##residues	787-1275	#label RES
##cross-references	GB:S62035; NID:9386046; PID:9386047	
CLASSIFICATION	#superfamily CDC25-type guanine nucleotide exchange activator homology; CDC24 homology; plexstrin repeat homology	
FEATURE		
240-426	#domain CDC24 homology #label CD24\	
1036-1272	#domain CDC25-type guanine nucleotide exchange activator homology #label SOS	
SUMMARY	#length 1275	#molecular-weight 145381 #checksum 4961
Query Match	4.6%;	Score 287; DB 2; Length 1275;
Best Local Similarity	25.6%;	Pred. No. 6,366;26;
Matches	40; Conservative	49; Mismatches 66; Indels 1; Gaps 1;
Db	1086 KHPNDISNLASIIINEDINARVSAIEKVAVADICRLHYNAVLEITSSMNSAIFR	1145
OY	656 RRPNEIQFWVTICLCSOISRRVLLKFKIFIAHCKEKNLMSFFAIWGLSNIAVSR	715
Db	1146 LKRTWLKVSQKOTALLDKLOKLVSSGGRKNLRREALKNDPCPVYLTOMYLTDLAFIEG	1205
OY	716 LALTWEKLPSEKFFFAEPESLMDPSRNIRARLYVAKIEPPLIFPMFLLIKDMFTHEG	775
Db	1206 TPNTEDGLVFSKMRMISHIIREIRFOQOTAYKIE	1241
OY	776 NKTFFIDN-IVNFEKMRMIANTATVATVRYRSQPFND	810
RESULT	6	
ENTRY	S28098	#type complete

TITLE	st66 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM	#formal_name Schizosaccharomyces pombe
DATE	17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
ACCESSIONS	S28098
REFERENCE	S28098
#authors	Hughes, D.A.; Fukui, Y.; Yamamoto, M.
#journal	Nature (1990) 344:355-357
#title	Nature (1990) 344:355-357
#cross-references	Homologous activators of ras in fission and budding yeast.
#accession	M0UD:90190870
#molecule-type	S28098
#residues	1-911 #label HUG
#cross-references	EMBL:X53254; NID:g5100; PID:g5101
GENETICS	
#gene	st66
CLASSIFICATION	#superfamily CDC25-type guanine nucleotide exchange activator homology
FEATURE	
659-694	#domain CDC25-type guanine nucleotide exchange activator homology #label S05
SUMMARY	#length 911 #molecular-weight 105185 #checksum 2084
Query Match	4.5%; Score 278; DB 2; Length 911;
Best Local Similarity	22.2%; Pred. No.1.89e-24;
Matches	57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;
Db	638 QLEDL5-LHNSPDDIYKDELVLPPRIKAKOLIEFOSFHSISRIOFTKIMDLN 696
Oy	584 EQPDLPLRPEQDEPFGVGFELMSNDLAVQMITYMELFNCVHELDLHYHFGHN 643
Db	697 RFSPEKETSTFYLSN-HLVNF-VTEITIOVEEPRRRTNLAYFIOVCDYLRNLNPSLF 754
Oy	644 -FKRTATNLDFLFRFNEIDQFMVVTETICCSQLSKRVOCLKFKIKIAHCKEYKNLSFF 702
Db	755 SIISALSSPIHRLRKRWANLNSKITLASELNLNLTFAKKNNSNRDOLENCYLPQVFL 814
Oy	703 AIVAGLSINIASRIALATWELKPSKKFAFEESLMDPSRNHRAARLVIAKLEPLIPFM 762
Db	815 GVVYTDLTFLKTGKDKDFQNMNINFDRTYTRILNKKFKQSGVGMFNPINEVOELNEV 874
Oy	763 PLIKDMFTFHGKGFPIIDLNVFEEKRMAMANTARYRIRYRQPPNPAQAQANKH-QDV 821
Db	875 ISREKNTNNI-YQKSLT 890
Oy	822 RSYVROLNVIDNQRTLS 838
RESULT	7
ENTRY	S25714 #type fragment
TITLE	son-of-sevenless-2 protein - mouse (fragment)
ORGANISM	#formal_name Mus musculus #common_name house mouse
DATE	13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-1999
ACCESSIONS	S25714
REFERENCE	S25714
#authors	Bowtell, D.; Fu, P.; Simon, M.; Senior, P.
#journal	Proc. Natl. Acad. Sci. U.S.A. (1992) 89:6511-6515
#title	Identification of murine homologues of the Drosophila Son of sevenless gene: potential activators of ras.
#cross-references	M0UD:92353328
#accession	S25714
#status	preliminary
#molecule-type	mRNA
#residues	1-1297 ##label BOW
#cross-references	EMBL:Z11664; NID:g941336; PID:g941337
CLASSIFICATION	#superfamily CDC25-type guanine nucleotide exchange activator homology; pleckstrin repeat homology
FEATURE	
405-507	#domain pleckstrin repeat homology #label PKV
739-982	#domain CDC25-type guanine nucleotide exchange activator homology #label S05
SUMMARY	#length 1297 #checksum 7302

Query Match	4.3%	Score 266:	DB 2:	Length 1297:
Best Local Similarity	24.9%	Pred. No. 1.67e-22:		
Matches	57:	Conservative	70:	Mismatches 92: Indels 10: Gaps 7:
Db	717	FESSPPVEWHISRTGQEFFDMLTHPIEIAQQLLESDDYRKQVPSLVGSWTKE	776	
Oy	586	FDSLPLPEQCGPTVGIVGTFFELMS--SKDLAQMITYDMELFNCVHELLYHFGGHN	643	
Db	777	KEINSPNLKRIKRTTNULTIMEFECYAEANFEERAVLSRIYELLQVODLNNGVLE	836	
Oy	644	FKKTTANIDFLRFRNFNIQFVWVTEICLSQSLRKVOLLKFKIKIAHCKEKNINSFE	703	
Db	837	IVSANSVSVYRLDHTFEALQER-KRIILD-DA-VESLDHFKKYLVKLKINSIPCVPFF	893	
Oy	704	IYMGISNIAVSRLALTMKEKLPSKFKRYAFESLMDPSKNH-RATRLYAKLEPPLIFPM	762	
Db	894	GIYLTNLIKTEEGNSDFLKRKGDLINFSKRRVAEITGEIOQYONQY	942	
Oy	763	PLLIKDMFTHEGKTFI--D-NLVFEKMRIMANTARTVRYYSQPF	807	
RESULT	8	A37488	#type complete	
ENTRY		Ras guanine nucleotide exchange factor son-of-sevenless (sos)		
TITLE		1 - human		
ORGANISM		#formal_name Homo sapiens #common_name man		
DATE		03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change 18-Mar-1997		
ACCESSIONS		A37488		
REFERENCE		A37488		
#authors		Charadlin, P.; Camonis, J.H.; Gale, N.W.; van Aelst, L.; Schlessinger, J.; Wigler, M.H.; Bar-Sagi, D.		
#journal		Science (1993) 260:1338-1343		
#title		Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2.		
#cross-references		MUID:93262494		
#accession		A37488		
#status		preliminary: not compared with conceptual translation		
#molecule_type		mRNA		
#residues		1-1333 ##label CHA		
#note		Sequence extracted from NCBI backbone (NCBIP:132148)		
CLASSIFICATION		#superfamily CDC25-type guanine nucleotide exchange activator		
FEATURE		homology: pleckstrin repeat homology		
442-544		#domain pleckstrin repeat homology #label PLK\		
776-1019		#domain CDC25-type guanine nucleotide exchange activator		
SUMMARY		length 1333 #molecular_weight 152473 #checksum 7142		
Query Match	4.1%	Score 255:	DB 2:	Length 1333:
Best Local Similarity	25.4%	Pred. No. 9.72e-21:		
Matches	58:	Conservative	65:	Mismatches 97: Indels 8: Gaps 6:
Db	754	FQSSPPTVEWHISKRGIEFFDMLTHPIEIAQQLLESDDYRAVOPSELVGSWTKE	813	
Oy	586	FDSLPLPEQCGPTVGIVGTFFELMS--SKDLAQMITYDMELFNCVHELLYHFGGHN	643	
Db	814	KEINSPNLKRIKRTTNULTIMEFECYAEANFEERAVLSRIYELLQVODLNNGVLE	873	
Oy	644	FKKTTANIDFLRFRNFNIQFVWVTEICLSQSLRKVOLLKFKIKIAHCKEKNINSFE	703	
Db	874	VVSAMNSSPVYRLDHTFEQIPSRQKILDEAHLSL-DHYK-YLAKLRISINPQVPPFG	931	
Oy	704	IYMGISNIAVSRLALTMKEKLPSKFKRYAFESLMDPSKNHATYAKLEPPLIFPM	763	
Db	932	IYLTNLIKTEEGNSDFLKRKGDLINFSKRRVAEITGEIOQYONQY	979	
Oy	764	PLLIKDMFTHEGKTFI--D-NLVFEKMRIMANTARTVRYYSQPF	807	
RESULT	9	S30356	#type complete	
ENTRY				

RESULT	ENTRY	8	9
TITLE		A37488	#type complete
ORGANISM		Ras guanine nucleotide exchange factor son-of-sevenless (sos)	
DATE		1 - human	
ACCESSIONS		#formal_name Homo sapiens #common_name man	
REFERENCE		03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change 18-Mar-1997	
#authors		A37488	
#journal		A37488	
#title		Chardin, P.; Camonis, J.H.; Gale, N.W.; van Aelst, L.; Schlessinger, J.; Wigler, M.H.; Bar-Sagi, D. Science (1993) 260:1338-1343	
#cross-references		Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2.	
#accession		MU01:93262494	
#status		A37488	
#molecule_type		preliminary: not compared with conceptual translation	
#residues		mRNA	
#note		1-1333 ##label CHA	
CLASSIFICATION		sequence extracted from NCBI backbone (NCBI:P:132148)	
FEATURE		#superfamily CDC25-type guanine nucleotide exchange activator	
442-544		homology: pleckstrin repeat homology	
776-1019		#domain CDC25-type guanine nucleotide exchange activator	
SUMMARY		#domain CDC25-type guanine nucleotide exchange activator	
		homology #label SOS	
		#length 1333 #molecular_weight 152473 #checksum 7142	
Query Match		4.1%; Score 255; DB 2; Length 1333;	
Best Local Similarity 25.4%;		Prod. No. 9.72e-21;	
Matches 58; Conservative 65; Mismatches 97; Indels 8; Gaps 6;			
Db		754 FQSSPPVEMHISNPGIEFDLLTHPIELIAQRLTLESDDLRAVQPSSELYGSVMTKD	813
Oy		586 FDSLPLPEQGGPGVGTGTFELMS--SKDLAYQMTIYDMEFNCVHELELYHTFGHNN	643
Db		814 KEINSPNLKIRITNTMLTMEFKCYIETDENLEERAAVYSRIETLLOVFOELNNPGYLE	873
Oy		644 FKKTATNIDLELRFNFIQWVYTECLCSQSLRKQQLKFKKIKIAHCKEYKNINSFPA	703
Db		874 VVSAMNSPPVRLDHPFEQIPSRQKILIEAHLELSE-DHYK-YLAKLRISINPCVPFEG	931
Oy		704 IVMGISINAVSRALITWEKLPSEKFKFYAFESLMDPSNNHRAIRLYIAKLEPPLIPMP	763
Db		932 IYLTNLITKEGNEPVYLKRHGEILINFSKRRKAAETIGELIQOYQNPY	979
Oy		764 LLIDMFTFHGNN-KTFI--DN-LVNFEMKRMIAANTARIVRRYSQPF	807
RESULT		S30356	#type complete
ENTRY			

Query Match	4.1%	Score 255:	DB 21:	Length 1366:
Best Local Similarity	25.4%	Pred. No. 9,72e-21:		
Matches	58;	Conservative	64;	Mismatches 98; Indels 8; Gaps 6;
Db	771	FOSSPPTVEMH1SRGHEIETDILLPHETIANQALLLESDDLVRAPVDSLGSVWTKED	830	
Oy	586	FDSLITPLPEQDEPPTGVTGTFELMS--SKDLAQMTYWMELFCVHELELITHTFGHN	643	
Db	831	KEINSNLKMKIRHNTNLTLMPEKCIYETENLEERAAVVSRIIEILQVFOELNPNVYLE	890	
Oy	644	EKKTTANLDELFRNRENIQFVWYVEILCSQLSKRQQLKFKFKIATNACEKYNLSFEA	703	
Db	891	VVSAMNSPVRILDTHEQDIFSRQKILIEPAHELSE-DHYKK-VYAKLRSINPCVPPFG	948	
Oy	704	IVAGLSINIVSRILATLWELKSKFKFYAEFESIMPSRNRHARVRLTYAKLEPPLIFMP	763	
Db	949	IVTLNLTKEEGSNPELRRHGKELINSKRRVAETGELGOONQY	996	
Oy	764	LILKDMFTFHEGN-KTFI--DN-LVNEFKRMATNARVRYRRSQPF	807	
RESULT	11	RGBC5	#type complete	
ENTRY			cell division control protein CDC25 - yeast (Saccharomyces cerevisiae)	
TITLE			protein I2142.6: protein YLR310C	
ALTERNATE_NAMES			#formal_name Saccharomyces cerevisiae	
ORGANISM			31-Mar-1988 #sequence_revision 31-Mar-1993 #text_change 26-Feb-1999	
ACCESSIONS			A26596; S51442; A23444; S43051; S47990	
REFERENCE			A26596	
#authors			Broek, D.; Toda, T.; Michaeli, T.; Levin, L.; Birchmeier, C. Zoller, M.; Powers, S.; Wiger, M.	
#journal			Cell (1987) 48:789-799	
#title			The S. cerevisiae CDC25 gene product regulates the RAS/adenylate cyclase pathway.	
#cross-references			WUID:87131091	
#accession			A26596	
#molecule_type			DNA	
#residues			1-1589 ##label BRO	
#cross-references			EMBL:M15458; NID:g171184; PID:g171185	
REFERENCE			S51437	
#authors			Pauley, A.	
#submission			submitted to the EMBL Data Library, November 1994	
#description			The sequence of S. cerevisiae cosmid I2142.	
#accession			S51442	
#molecule_type			DNA	
#residues			1-1589 ##label PAU	
#cross-references			EMBL:U17247; NID:g577216; PID:g577222; MIPS:YLR310C	
REFERENCE			A23444	
#authors			Cemonis, J.H.; Kalejine, M.; Gondre, B.; Garreau, H.; Boy-Marcotte, E.; Jaquet, M.	
#journal			EMBO J. (1986) 5:375-380	
#title			Characterization, cloning and sequence analysis of the CDC25 gene which controls the cyclic AMP level of Saccharomyces cerevisiae.	
#cross-references			WUID:86220116	
#accession			A23444	
#molecule_type			DNA	
#residues			1-496 'Y', 498-953 'LSYINMLSR', 964-1589 ##label CAM	
#cross-references			EMBL:X03579; NID:g3483; PID:g3484	
REFERENCE			S43051	
#authors			Daniel, J.H.	
#journal			Curr. Genet. (1986) 10:879-885	
#title			The CDC25 "Start" gene of Saccharomyces cerevisiae: sequencing of the active C-terminal fragment and regional homologues with rhodopsin and cytochrome P450.	
#cross-references			WUID:88194639	
#accession			S43051	
#molecule_type			DNA	
#residues			877-1589 ##label DAN	
#cross-references			EMBL:X03579	

#gene SGD:CDCC25; CTN1  
#cross-references SGD:S0004301; MIPS:YLR310c  
#map position 12R  
#description positive control of level of cellular cAMP at the stage at which the cell division cycle is triggered  
#classification #superfamily budding yeast CDC25; CDC25-type guanine nucleotide exchange activator homology; SH3 homology  
#keywords cell cycle control; transmembrane protein  
#feature  
65-123 #domain SH3 homology #label SH3\  
1301-1542 #domain CDC25-type guanine nucleotide exchange activator homology #label SOS  
#summary #length 1589 #molecular-weight 179090 #checksum 7509

Query Match 4.1%; Score 255; DB 1; Length 1589;  
Best local Similarity 25.5%; Pred. No. 9,72e-21;  
Matches 71; Conservative 77; Mismatches 117; Indels 13; Gaps 10;

Db 1228 VSEKIPGAEIDLQKINEKLINEKEPDPKQODVSAYVQTTRKDKSPI-HMS-SSL 1285  
QY 537 VADKLGSCEGLI-IV-KMSSGGEKVLKPNV-SVFTTLTLINGRLFPACRQFDSLPL 592  
Db 1286 PSSASSAFPLKILKILDLIDPYATQQLVLEHDLVLRIMECDLRAGTKYCNMGSP 1345  
QY 593 PEQGPVGVGVTELMSSKDLAY--OMTYDDELNCVHELELIYHTRHNFKKT-IA 649  
Db 1346 NITKFINANTLINFVSHITVQADYKTRSKLTQYVTAHQCKELNFSMTAIVSALY 1405  
QY 650 NLDFLFRFHEIQWVWTEICLSQSLKRVQLKFKIAHCKEYKLNLSFPAIWMGLS 709  
Db 1406 SSPIYRKLKTKWDVSTESKDLKLNLMNSKRNFEVYRELLSVNDVACVPEGYLSD 1465  
QY 710 NIAVSRLATLWELKPSKFKFYEFESLMDPSRNHRAVR-LTYAKLEPPLIPMLLIK 768  
Db 1466 LTFEFGNPDPFLHNSNTINIEFSKRTKIANIVEETISFK 1503  
QY 769 MTFHEGKTFPI-D-NLVNEKRMITANRATRYRYR 803

RESULT 12  
ENTRY A41216 #type complete  
TITLE guanine nucleotide exchange activator son-of-sevenless - fruit fly (Drosophila melanogaster)  
ORGANISM #formal\_name Drosophila melanogaster  
DATE 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 24-Sep-1998  
ACCESSIONS A41216  
REFERENCE A41216  
#authors Simon, M.A.; Bowtell, D.D.L.; Dodson, G.S.; Laverly, T.R.; Rubin, G.M.  
#journal Cell (1991) 67:701-716  
#title Rasi and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase.  
#cross-references MUID:92034991  
#accession A41216  
#status preliminary  
#molecule\_type mRNA  
#residues 1-1596 ##label SIM  
#cross-references GB:M77501; NID:g158470; PID:g158471  
GENETICS  
#gene FlyBase:Sos  
#cross-references FlyBase:FBgn0001965  
CLASSIFICATION #superfamily CDC25-type guanine nucleotide exchange activator homology; pleckstrin repeat homology  
FEATURE  
#feature 479-586 #domain pleckstrin repeat homology #label PLK\  
825-1066 #domain CDC25-type guanine nucleotide exchange activator homology #label SOS  
#summary #length 1596 #molecular-weight 177933 #checksum 4040  
Query Match 4.0%; Score 251; DB 2; Length 1596;

Best local Similarity 24.4%; Pred. No. 4,22e-20;  
Matches 48; Conservative 59; Mismatches 86; Indels 4; Gaps 3;

Db 832 ELARQLILEFEMTKNKPSELVSGSPWTKDKDEKSNLKKHHTNVRWIEKSTEA 891  
QY 613 DLAYOMTYDDELNCVHELELIYHTRHNFKKTANLDFLRFEHIFWVTEICLC 672  
Db 892 ENYERIALIMGRAIEVMVMELNNGIISIVAMGTASYRLRWFQGLPERYKFELE 951  
QY 673 SOLSKRVQLKFKFKIAHCKEYKLNLSFPAIWMGLSINAVSRALTWELKPSKFKFYA 732  
Db 952 ECRELSD-DHLK-YOERLRISINPCVPEFGRYLTNILLHEGNDLILINESKRR 1009  
QY 733 EFESLMDPSRNHRAVRILVAKLEPPLIPFMPLLIKDMTFHEGKTFIDN-LVNEFKMR 790  
Db 1010 KVAETIGEIOQYONPY 1026  
QY 791 MIANTARTVRYRSQPF 807

RESULT 13  
ENTRY S64758 #type complete  
TITLE SCD25 protein (version 2) - yeast (Saccharomyces cerevisiae)  
ALTERNATE\_NAMES protein L1309; protein YIL016w  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 01-Aug-1995 #sequence\_revision 24-May-1996 #text\_change 06-Feb-1998  
ACCESSIONS S64758; S64764; S69390; S70559  
REFERENCE S64743  
#authors Miosga, T.; Zimmermann, F.K.  
#submission submitted to the Protein Sequence Database, May 1996  
#accession S64758  
#molecule\_type DNA  
#residues 1-1048 ##label MIO  
#cross-references EMBL:T73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YIL016w  
#experimental\_source strain S288C  
REFERENCE S64761  
#authors Goffeau, A.; Purnelle, B.  
#submission submitted to the Protein Sequence Database, May 1996  
#accession S64764  
#molecule\_type DNA  
#residues 1-1048 ##label GOF  
#cross-references EMBL:T73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YIL016w  
#experimental\_source strain S288C  
REFERENCE S69380  
#authors Purnelle, B.; Goffeau, A.  
#submission submitted to the EMBL Data Library, April 1996  
#description The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14 open reading frames among which HSP104, SSA2, SPA2, KNS1, DPS1/APS, SDC25, a new member of the serpauperins family and a new ABC transporter homologous to the human multidrug resistance protein.  
#accession S69390  
#molecule\_type DNA  
#residues 1-1048 ##label PUR  
#cross-references EMBL:X97560; NID:g1297003; PID:e238680; PID:g1297014  
REFERENCE S70557  
#authors Miosga, T.; Zimmermann, F.K.  
#journal Yeast (1996) 12:693-708  
#title Sequence analysis of the CEN12 region of chromosome XII including an open reading frame homologous to the human cystic fibrosis transmembrane conductance regulator protein CFTR.  
#cross-references MUID:96405918  
#accession S70559  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-1048 ##label MIW  
#cross-references EMBL:X91488; NID:g1495203; PID:g199006; PID:g1495207  
#note the nucleotide sequence was submitted to the EMBL Data

CLASSIFICATION	#superfamily CDC25-type guanine nucleotide exchange activator homology
FEATURE	
946-1197	#domain CDC25-type guanine nucleotide exchange activator homology #label S05
SUMMARY	#length 1250 #molecular-weight 144779 #checksum 1783
Query Match	3.8%; Score 236; DB 2; Length 1250;
Best Local Similarity	3.0%; Pred. No. 9,82e-18;
Matches	35; Conservative 26; Mismatches 44; Indels 1; Gaps 1;
Db	1018 SKRAKLHFFIAEYCRKFNFSMTDILASYSPIREKTKWQAVIPQTFDLQSLN 1077
QY	676 SKRYDLKKFKIKIAHCKEYKKNNSFFAIVMLSIINAISRLATWEXLPKFKKFAEPE 735
Db	1078 KLMDEPKNFINRNELKSLSACVCFEYVSLDTFTFDSGNPYL 1123
QY	736 SLMDPSRHHRAVRLTVAKLER-PLIFPMVLLIKDMFTTHBENKRFI 780
RESULT	15
ENTRY	OKGARI
TITLE	#type complete protein kinase (EC 2.7.1.37), cAMP-dependent, type I regulatory chain - California sea hare
ORGANISM	#formal_name Aplysia californica #common_name California sea hare
DATE	31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 05-Sep-1997
ACCESSIONS	JH0590
REFERENCE	JH0590
#authors	Bergold, P.J.; Beushausen, S.A.; Sacktor, T.C.; Cheley, S.; Bayley, H.; Schwartz, J.H.
#journal	Neuron (1992) 8:387-397
#title	A regulatory subunit of the cAMP-dependent protein kinase down-regulated in aplysia sensory neurons during long-term sensitization.
#cross-references	MUID:92135429
#accession	JH0590
#molecule_type	mRNA
#residues	1-378 #label BER
#cross-references	EMBL:X62382; NID:95587; PID:95588
#note	this protein is identified as the M4 isoform
COMMENT	The inactive form of the enzyme is composed of two regulatory chains and two catalytic chains. Activation by cAMP produces two active catalytic monomers and a regulatory dimer that binds four cAMP molecules.
COMMENT	Type I regulatory chains contain a high-affinity binding site for MgATP.
COMMENT	The phosphophorylation site binds to the substrate-binding region of the catalytic chain but is not phosphorylated. The physiological significance of phosphorylations by other kinases is unclear
CLASSIFICATION	#superfamily cAMP-dependent protein kinase regulatory chain; cAMP receptor protein cyclic nucleotide-binding domain
KEYWORDS	acetylated amino end; ATP binding; cAMP binding; duplication; heterotetramer; homodimer; phosphoprotein; phosphotransferase
FEATURE	
2-378	#product protein kinase, cAMP-dependent, type I-beta regulatory chain #status predicted #label M4V
2-133	#domain protein interaction #label DMV
93-97	#region cAMP-dependent phosphorylation motif
134-251	#domain cAMP receptor protein cyclic nucleotide-binding domain homology #label CA1V
252-378	#domain cAMP receptor protein cyclic nucleotide-binding domain homology #label CA2V
2	#modified site acetylated amino end (ala) (in mature form) #status predicted
199,208	#binding_site cAMP (Glu, Arg) #status predicted
323,332	#binding_site cAMP (Glu, Arg) #status predicted
SUMMARY	#length 378 #molecular-weight 42737 #checksum 2977

Sun Apr 23 09:32:16 2000

US-09-422-999-18.rpr

Page 8

	Query Match	3.5%	Score 216:	DB 1:	length 378;
	Best Local Similarity	32.7%;	Pred. No.	1.21e-14;	
	Matches	33;	Conservative	26;	Mismatches 39; Indels 3; Gaps 3;
Dd	135 FSHLDNERSDIDIMAF-PVHRHAGEVIIOGDGDFYVIDGEVDYDV-NNNVHTSIG	192			
	::::: ::: :   :   :   :   :   :   :   :   :   :   :				
Oy	195 LSHSTIVKRELAVIIFESHAKKGTIVLFQNGEGSTISWYIIILKGSNVNIYIKGVCTIH	254			
Dd	193 EGGSFELALIVTIPRAATVKAKTD-KLWDIDRDSRYLL	232			
	:::          : :   :   :   :   :   :   :				
Oy	255 EGGDFGRLALYNDAAPRAASIVLRDNCHFLNVKDREPNRI	295			

Search completed: Sat Apr 22 13:19:19 2000  
Job time : 39 secs.



\*\*\*\*\*  
MWSEH  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mparch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 22 13:23:13 2000; MasPar time 14.86 Seconds  
Tabular output not generated. 759.883 Million cell updates/sec

Title: >US-09-422-999-18  
Description: (1-849) from US09422999.pep  
Perfect Score: 6222  
Sequence: 1 MVLKRALVTMTGFLTRTHL.....VIDNQRLSOMSHRLPRRP 849

Scoring table: PAM 150  
Gap 11

Searched: 134018 seqs, 13297625 residues

Post-Processing: Minimum Match 08  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:6\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 36.052; Variance 179.662; scale 0.201

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	287	4.6	489	1	US-08-318-Sequence 4, Applicatio	1.23e-12
2	287	4.6	666	1	US-08-318-Sequence 3, Applicatio	1.23e-12
3	287	4.6	814	1	US-08-318-Sequence 2, Applicatio	1.23e-12
4	276	4.4	402	2	US-08-290-Sequence 14, Applicati	8.79e-12
5	276	4.4	652	1	US-08-318-Sequence 8, Applicatio	8.79e-12
6	266	4.3	423	2	US-08-290-Sequence 11, Applicati	5.22e-11
7	266	4.3	1297	2	US-08-290-Sequence 4, Applicatio	5.22e-11
8	256	4.1	354	1	US-08-318-Sequence 6, Applicatio	3.07e-10
9	255	4.1	423	2	US-08-290-Sequence 10, Applicati	3.07e-10
10	255	4.1	426	2	US-08-290-Sequence 12, Applicati	3.67e-10
11	255	4.1	1319	2	US-08-290-Sequence 2, Applicatio	3.67e-10
12	255	4.1	1336	2	US-08-290-Sequence 6, Applicatio	3.67e-10
13	251	4.0	430	2	US-08-290-Sequence 9, Applicatio	7.43e-10
14	251	4.0	1572	2	US-08-290-Sequence 5, Applicatio	7.43e-10
15	236	3.8	418	2	US-08-290-Sequence 13, Applicati	1.03e-08
16	191	3.1	362	2	US-08-290-Sequence 15, Applicati	1.39e-05
17	143	2.3	852	4	PCT-US95-0Sequence 5, Applicatio	6.37e-02
18	143	2.3	852	4	US-08-408-Sequence 5, Applicatio	6.37e-02
19	133	2.1	768	4	PCT-US95-0Sequence 2, Applicatio	3.07e-01
20	133	2.1	768	2	US-08-408-Sequence 2, Applicatio	3.07e-01
21	122	2.0	528	1	US-08-363-Sequence 14, Applicati	1.67e+00
22	122	2.0	530	1	US-08-363-Sequence 3, Applicatio	1.67e+00
23	122	2.0	530	1	US-08-363-Sequence 10, Applicati	1.67e+00

24	122	2.0	530	1	US-08-363-Sequence 9, Applicatio	1.67e+00
25	122	2.0	530	1	US-08-363-Sequence 8, Applicatio	1.67e+00
26	122	2.0	530	1	US-08-363-Sequence 2, Applicatio	1.67e+00
27	120	1.9	210	2	US-08-892-Sequence 2, Applicatio	2.26e+00
28	116	1.9	451	1	US-08-625-Sequence 4, Applicatio	4.13e+00
29	119	1.9	1159	2	US-08-956-Sequence 13, Applicati	2.63e+00
30	106	1.6	888	2	US-08-956-Sequence 2, Applicatio	1.81e+01
31	98	1.6	178	1	US-07-929-Sequence 4, Applicatio	5.68e+01
32	98	1.6	184	2	US-08-945-Sequence 10, Applicati	5.68e+01
33	97	1.6	442	1	US-08-363-Sequence 11, Applicati	6.53e+01
34	97	1.6	445	1	US-08-363-Sequence 5, Applicatio	6.53e+01
35	99	1.6	480	1	US-07-803-Sequence 2, Applicatio	4.93e+01
36	101	1.6	543	2	US-08-922-Sequence 10, Applicati	3.71e+01
37	97	1.6	858	1	US-07-712-Sequence 2, Applicatio	6.53e+01
38	102	1.6	1167	1	US-08-485-Sequence 6, Applicatio	3.21e+01
39	102	1.6	1167	2	US-09-184-Sequence 6, Applicatio	3.21e+01
40	102	1.6	1167	2	US-08-590-Sequence 6, Applicatio	3.21e+01
41	102	1.6	1168	1	US-08-620-Sequence 9, Applicatio	3.21e+01
42	98	1.6	2319	1	US-08-212-Sequence 8, Applicatio	5.68e+01
43	98	1.6	2319	4	PCT-US94-1Sequence 6, Applicatio	5.68e+01
44	98	1.6	2319	1	US-08-474-Sequence 6, Applicatio	5.68e+01
45	98	1.6	2319	2	US-08-670-Sequence 6, Applicatio	5.68e+01

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	489	AA.
ID	US-08-318-831-4				
XX	xxxxxx				
AC					
XX					
DT					
XX					
DE	Sequence 4, Application US/08318831				
XX					
CC	Sequence 4, Application US/08318831				
CC	Patent No. 5656595				
CC	GENERAL INFORMATION:				
CC	APPLICANT: Schweighofer, Fabien				
CC	INVENTOR: Tocque, Bruno				
CC	TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR				
CC	TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SAID PEPTID				
CC	TITLE OF INVENTION: PREPARATION AND UTILIZATION				
CC	NUMBER OF SEQUENCES: 12				
CC	CORRESPONDENCE ADDRESS:				
CC	ADDRESS: Rhone-Poulenc Rorer Inc.				
CC	STREET: 500 Arcola Road, 3C43				
CC	CITY: Collegeville				
CC	STATE: PA				
CC	COUNTRY: USA				
CC	ZIP: 19426				
CC	COMPUTER READABLE FORM:				
CC	MEDIUM TYPE: Floppy disk				
CC	COMPUTER: Macintosh				
CC	OPERATING SYSTEM: System 7.1				
CC	SOFTWARE: Word 5.1 (Epo Patentin)				
CC	CURRENT APPLICATION DATA:				
CC	APPLICATION NUMBER: US/08/318, 831				
CC	FILING DATE: 19 October 1994				
CC	CLASSIFICATION: 435				
CC	PRIOR APPLICATION DATA:				
CC	APPLICATION NUMBER: FR92/04827				
CC	FILING DATE: 21-Apr-1992				
CC	ATTORNEY/AGENT INFORMATION:				
CC	NAME: Smith, Julie K.				
CC	REGISTRATION NUMBER: P-38, 619				
CC	REFERENCE/DOCKET NUMBER: ST92033-US				
CC	TELEPHONE: (610) 454-3839				
CC	TELEFAX: (610) 454-3808				
CC	INFORMATION FOR SEQ ID NO: 4:				
CC	SEQUENCE CHARACTERISTICS:				
CC	LENGTH: 489 amino acids				

625 KHFNDISNLAISEIRNEDINARVSAIEKVAADICRCLHNNYNAVEITSSMNRSAIFR 684

QY 656 RRENIOFWVTEICLSQLSKRVQLKKFKIAHCKEYKNNNSPFAIYMGISNTAVSR 715  
DB 685 LKKTWIKSKQKRALIDKLOKIVSSEGRFNKLRLPALKNCDDPCVPYIGMTLDLAFIEEG 744  
QY 716 LALTWEKLPSEKKEKYAFESLMDPSRNRHRAVRLTVAKLEPPLIPFMPFLIKMDITTHBG 775  
DB 745 TPVYTEDGLVNSFKRMISHIREIROPQOTAKIE 780  
QY 776 NKTFLDN-LVNEFKRMIAHTAKRYRYSQFPND 810

RESULT 4  
ID US-08-290-731C-14 STANDARD; PRT; 402 AA.  
XX xxxxxx  
XX

DE Sequence 14, Application US/08290731C  
XX  
CC Sequence 14, Application US/08290731C  
CC Patent No. 5843646  
CC GENERAL INFORMATION:  
CC APPLICANT: BOWTELL, David Douglas Lawrence  
CC TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE  
CC TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,  
CC TITLE OF INVENTION: AND MSOS POLYPEPTIDES  
CC NUMBER OF SEQUENCES: 15  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
CC STREET: 2100 PENNSYLVANIA AVENUE, N.W.  
CC CITY: WASHINGTON  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20037  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/290,731C  
CC FILING DATE: 17-OCT-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/AU93/00068  
CC FILING DATE: 17-FEB-1993  
CC PRIOR APPLICATION DATA: PLO921/92  
CC APPLICATION NUMBER: PLO921/92  
CC FILING DATE: 17-FEB-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: KIT, Gordon  
CC REGISTRATION NUMBER: 30,764  
CC REFERENCE/DOCKET NUMBER: O-36066  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 293-7060  
CC TELEFAX: (202) 293-7860  
CC INFORMATION FOR SEQ ID NO: 14:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 402 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 402 AA; 844812 CN;

Query Match 4.4%; Score 276; DB 2; Length 402;  
Best Local Similarity 22.2%; Pred. No. 8,79e-12;  
Matches 57; Conservative 75; Mismatches 119; Indels 6; Gaps 6;

DB 146 QLEEDLS-LHNSPDPIYDELVLPPREIAKOLCLIFPSFSHISRIOPFKINDEIN 204  
QY 584 EOPDSILPLPEOBPGVIGVTFELMSSKDLAYQWTIYDWELFNCVHELELIYHTEGRHN 643

DB 205 RFSPEKSTFEVLSN-HLVNF-VTEFIVQEEPRRRINVLAFYIOVDYRELNFASLF 262  
QY 644 -FKRTYANDLFLRRNEIOFWVTEICLSQLSKRVQLKKFKIAHCKEYKNNNSFF 702  
DB 263 SIISALNSPPIHRLKRTVMANLSKTLASPELNNLTERRKNSVNRQLENCVPCVPFL 322  
QY 703 AIYWGLSNINAVSLATWELKPSKKEKYAFESLMDPSRNRHRAVRLTVAKLEPPLIPFM 762  
DB 323 GYVFTDITFLTKGNKDNFCQNMINDKRTKVRILNEIKFQSGVGMNPINVOELINEV 382  
QY 763 PLIKMDTFTHESGKTFIDNLVNEFKRMIAHTAKRYRYSQFPNDAAQANKNH-ODV 821  
DB 383 ISRENTNNI-YORSLT 398  
QY 822 RSYROLNVIQDNQRTLS 838

RESULT 5  
ID US-08-318-831-8 STANDARD; PRT; 652 AA.  
XX xxxxxx  
XX

DE Sequence 8, Application US/08318831  
XX  
CC Sequence 8, Application US/08318831  
CC Patent No. 5656595  
CC GENERAL INFORMATION:  
CC APPLICANT: Schweigboffer, Fabien  
CC APPLICANT: Tocque, Bruno  
CC TITLE OF INVENTION: PEPTIDES HAVING A GDP EXCHANGE FACTOR  
CC TITLE OF INVENTION: ACTIVITY, NOCLIC ACID SEQUENCES CODING FOR SAID PEPTID  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Rhone-Poulenc Rorer Inc.  
CC STREET: 500 Arcola Road, 3043  
CC CITY: Collegeville  
CC STATE: PA  
CC COUNTRY: USA  
CC ZIP: 19426  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: Macintosh  
CC OPERATING SYSTEM: System 7.1  
CC SOFTWARE: Word 5.1 (EPO PatentIn)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/318,831  
CC FILING DATE: 19 October 1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: FR97/04827  
CC FILING DATE: 21-APR-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Smith, Julie K.  
CC REGISTRATION NUMBER: P-38,619  
CC REFERENCE/DOCKET NUMBER: ST92033-US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (610) 454-3839  
CC TELEFAX: (610) 454-3808  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 652 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC SEQUENCE 652 AA; 74235 MW; 2320730 CN;

Query Match 4.4%; Score 276; DB 1; Length 652;  
Best Local Similarity 25.8%; Pred. No. 8,79e-12;  
Matches 59; Conservative 71; Mismatches 89; Indels 10; Gaps 7;



[illegible]

Query Match	4.1%:	Score 256:	DB 1:	Length 364:
Best Local Similarity	25.4%:	Pred. No. 3.07e-10:		
Matches	58:	Conservative	65:	Mismatches 97: Indels 8: Gaps 6:
Db	22	FOSSPPVEMWHISPPGHIEFTDILLPHIEIAROLTLLSDLYRAVOPSDLVGSVMTKED	81	
Qy	586	FDLSLPLPEDEOGPIVGVGFTEFLMS--SKDLAVQMTIYDMEFLNFCVHELILYHTTGRHN	643	
Db	82	KEINSPLULKAIKRTTLLTLMEFEKCIYETENLEERRAVVSRRIIEILOVFOELNNGVLE	141	
Qy	644	EKKTTAMIDLFRFRFNIOFWVYVEICLSQSLSKRVOLKFKIKIAHCKEYENLNSFFA	703	
Db	142	VVSAMNSPPYRLDHTDEQIOPROKILTEEAHKESE-DHYKK-VIAKLRSINPCVDFEG	199	
Qy	704	IWGLSINIAVSRLATLWEKLPSKKKRYAEFESIMDPSRNRHRAVRLTVAKLEPLIPMP	763	
Db	200	LYLNNIKTEGNEPEVLKRHGKELINFSKRRKVAETIGELIQOYONOPY	247	
Qy	764	LLIDMFTHEGN-KTFI--DN-LVNEKRMIAANTARTVRYRSQPF	807	
RESULT	9	STANDARD;	PRT:	423 AA.
ID	US-08-290-731C-10			
AC	xxxxxx			
XX				
DT				
DE	Sequence 10, Application US/08290731C			
XX				
CC	Sequence 10, Application US/08290731C			
CC	Patent No. 5843646			
CC	GENERAL INFORMATION:			
CC	APPLICANT: BOWTELL, David Douglas Lawrence			
CC	TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE			
CC	TITLE OF INVENTION: SON OF SEVENTHLESS (MSOS) GENE,			
CC	TITLE OF INVENTION: AND MSOS POLYPEPTIDES			
CC	NUMBER OF SEQUENCES: 15			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: SUGHRUE, MTON, ZINN, MACPEAK & SEAS			
CC	STREET: 2100 PENNSYLVANIA AVENUE, N.W.			
CC	CITY: WASHINGTON			
CC	STATE: D.C.			
CC	COUNTRY: USA			
CC	ZIP: 20037			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Floppy disk			
CC	COMPUTER: IBM PC compatible			
CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	SOFTWARE: Patentin Release #1.0, Version #1.25			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/290,731C			
CC	FILING DATE: 17-OCT-1994			
CC	CLASSIFICATION: 435			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER: PCT/UA93/00068			
CC	FILING DATE: 17-FEB-1993			
CC	PRIOR APPLICATION DATA:			
CC	APPLICATION NUMBER: PL0921/92			
CC	FILING DATE: 17-FEB-1992			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: KIT, Gordon			
CC	REGISTRATION NUMBER: 30,764			
CC	REFERENCE/DOCKET NUMBER: Q-36066			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: (202) 293-7060			
CC	TELEFAX: (202) 293-7860			
CC	TELEX: 6491103			
CC	INFORMATION FOR SEQ. ID NO: 10:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 423 amino acids			
CC	TYPE: amino acid			

CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 426 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 426 AA; 49080 MW; 1001703 CN;
CC	
CC	Query Match 4.1%; Score 255; DB 2; Length 426;
CC	Best Local Similarity 25.5%; Pred.No.3,67e-10;
CC	Matches 71; Conservative 77; Mismatches 117; Indels 13; Gaps
CC	
CC	Db 108 VSEKIPGAEIDLQKINEXLINEKEPEPDQODSVSAVYQTTKRDNKSPI-HMS-SSL 1655
CC	QY 537 VADKIGSGEGLI-IV-KMSGGGERKVVYKLPNDV-SVFTTLITINGRLFACPREQFDSLTPL 5922
CC	Db 166 PSSASAEFRLLKKLKLIDIDPYTATOLTVLEHDLVLRITWFEGLDRAMCTKCNMGSGP 2253
CC	QY 593 PEQEPPIYGVGTPELMSKSLAY-QWTITDWEFLMCAHELELITFTGRHNPKT-TA 6499
CC	Db 226 NITKFIANANTLTNEVSHTVIKQADVTKRSKLTQYFYVAOHCKELNFFSSMTAISALY 2853
CC	QY 650 NLDLPLRFNEIOWVYVEICGSLSRVGLKKFIKIAHCKEYKNTLMSFAIVGLS 7099
CC	Db 286 SSPIRLKRTMDLVSTESKDLKLNLDKSKRNFVYKREILNSYTDVACVPPFGVYLS 3455
CC	QY 710 NIAVSRLLLTWEKLEPSKKKKYFAFEFESLMPSSRNHRYR-LTVAKLEPPLIPMPPLIKD 7688
CC	Db 346 LTFEFGNPDFLHNSNTIINSKRTKIANIYEELISPK 383
CC	QY 769 MTFHEGKTFI-D-NLVNEFKRMRIANKARTVRYR 803
CC	
CC	RESULT 11
CC	ID US-08-290-731C-2 STANDARD; PRT; 1319 AA.
CC	XX
CC	DT xxxxxx
CC	XX
CC	Sequence 2, Application US/08290731C
CC	XX
CC	Sequence 2, Application US/08290731C
CC	DE Patent No. 5843646
CC	XX
CC	GENERAL INFORMATION:
CC	APPLICANT: BOWTELL, David Douglas Lawrence
CC	TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
CC	TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
CC	TITLE OF INVENTION: AND MSOS POLYPEPTIDES
CC	NUMBER OF SEQUENCES: 15
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: SUGHRUE, MTON, ZINN, MACPEAK & SEAS
CC	STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CC	CITY: WASHINGTON
CC	STATE: D.C.
CC	COUNTRY: USA
CC	ZIP: 20037
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/290,731C
CC	FILING DATE: 17-OCT-1994
CC	CLASSIFICATION: 435
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/AU93/00068
CC	FILING DATE: 17-FEB-1993
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: PL0921/92
CC	FILING DATE: 17-FEB-1992
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: KIT, Gordon

```
CC FILING DATE: 17-FEB-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KIT, Gordon
CC REGISTRATION NUMBER: 30,764
CC REFERENCE/DOCKET NUMBER: Q-36066
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 293-7060
CC TELEFAX: (202) 293-7860
CC TELEX: 6491103
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1336 amino acids
CC TYPE: amino acids
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1336 AA; 152734 MW; 9610574 CN;

Query Match          4.1%; Score 255; DB 2; Length 1336;
Best Local Similarity 25.4%; Pred. No. 3.67e-10;
Matches    58; Conservative   64; Mismatches 98; Indels    8; Gaps     6

Db      771 FOSSPVEWMISRGHIEFDLLHPHEIAROLTLLESDLRAVOPSELVGSWTKED 830
Qy      566 FDSLPLPEGSGPIYGVITGFELMS--SKDLAYMTIDWLFNFCVHELILYTFCGRHN 643
Db      831 KEINSPLLKMRHTTNTLTFEKCIVETENLEERVAVSRRIIELLOVFDELNNFGYLE 890
Qy      644 EFKTTANLDLRFRNEFIOWFWVTIEICLSQLRSKVOLLKKFIKAIAHKCKEYMKNINSEFA 703
Db      891 VVSANNSSPYVRLDHTFPQIPROKLIKLEAHELSE-DHYKK-VLAKRSLNPCVPFFG 948
Qy      704 IVMGISNAIVSRALATWRLKPSPRKRYFAEEESLMDDPSRNHRAYRITYAKLEPLLPMP 763
Db      949 IVLTNIILTEEGNPVELRRHGKELINEFSKRRRVAEITGEIOYONOPY 996
Qy      764 LLIRMDTHTEGN-KTFI--DN-LVNFEKMIMANTARTATVYYTSQP 807

RESULT 13
ID US-08-290-731C-9 STANDARD; PRT; 430 AA.
XX xxxxxx
XX DE
XX Sequence 9, Application US/08290731C
XX CC
XX Patent No. 5843646
XX CC
XX GENERAL INFORMATION:
XX APPLICANT: BOWTELL, David Douglas Lawrence
XX TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
XX TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
XX TITLE OF INVENTION: AND MSOS POLYPEPTIDES
XX NUMBER OF SEQUENCES: 15
XX CORRESPONDENCE ADDRESSES:
XX ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS
XX STREET: 2100 PENNSYLVANIA AVENUE, N.W.
XX CITY: WASHINGTON
XX STATE: D.C.
XX COUNTRY: USA
XX ZIP: 20037
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX OPERATING SYSTEM: IBM PC compatible
XX SOFTWARE: PatentIn Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/290,731C
XX FILING DATE: 17-OCT-1994
XX CLASSIFICATION: 435
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: PCT/AU93/00068
```

CC FILING DATE: 17-OCT-1994

CC SOFTWARE: PatentIn Release #1.0, Version #1.25



```

CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/290,731C
CC      FILING DATE: 17-OCT-1994
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/AU93/00068
CC      FILING DATE: 17-FEB-1993
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: PLO921/92
CC      FILING DATE: 17-FEB-1992
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: KIT, Gordon
CC      REGISTRATION NUMBER: 30,764
CC      REFERENCE/DOCKET NUMBER: Q-36066
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (202) 293-7060
CC      TELEFAX: (202) 293-7860
CC      TELEX: 6491103
CC      INFORMATION FOR SEQ ID NO: 13:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 418 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
SQ      SEQUENCE 418 AA; 48715 MW; 949499 CN;

Query Match      3.8%; Score 236; DB 2; Length 418;
Best Local Similarity 33.0%; Pred. No. 1.03e-08;
Matches 35; Conservative 26; Mismatches 44; Indels 1; Gaps 1;

Db      235 SKRAKLSHPTFAEYGRKFNFSMDIISALYSPPIYRLEKRWQAVIPQTRDLQSLN 294
Qy      676 SKRVQLKKFTKIAHCKEYKRLNSFEATYVGLSNIAVSRLATLWETKLPKFKFTAEFE 735

Db      295 KIMDPKKNFINRYNELKSLHSPCVPEFGVYLSDLFTDSGNPDYL 340
Qy      736 SIMDPSRNHRAVRLTVAKLEP-PLIPFMPLLIKDMFTHEGKKTFFI 780
  
```

Search completed: Sat Apr 22 13:23:31 2000  
 Job time : 18 secs.





(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msearch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 22 13:17:33 2000; Maspar time 44.93 Seconds

Tabular output not generated. 447.554 Million cell updates/sec

Title: >US-09-422-999-18  
Description: (1-849) from US09422999.dep  
Perfect Score: 6222  
Sequence: 1 MVLKRALMTGFLTRRTL.....VINQRTLSQMSHRLPRRP 849

Scoring table: PAM 150  
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseq

Statistics: Mean 38.706; Variance 183.913; scale 0.210

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	370	5.9	1077	1	R91597	C3G protein.	1.53e-19
2	287	4.6	489	1	R50991	Peptide which modulate	9.88e-13
3	287	4.6	666	1	R50990	Peptide which modulate	9.88e-13
4	287	4.6	814	1	R43578	Peptide which modulate	9.88e-13
5	276	4.4	652	1	R43580	Peptide which modulate	7.60e-12
6	266	4.3	1297	1	R47048	Mammalian son of seven	4.81e-11
7	262	4.2	1297	1	R84639	mSOS2 protein.	1.00e-10
8	256	4.1	364	1	R45379	Peptide which modulate	3.02e-10
9	255	4.1	1319	1	R47043	Mammalian son of seven	3.63e-10
10	255	4.1	1336	1	R84638	mSOS1 protein.	3.63e-10
11	201	3.2	380	1	W54350	CAMP-dependent protein	5.94e-06
12	180	2.9	609	1	W87995	An alternatively splic	2.29e-04
13	180	2.9	728	1	W87994	A human MCG7 protein.	2.29e-04
14	134	2.2	932	1	W85040	PKG-green fluorescent	4.81e-01
15	134	2.2	933	1	W85039	Green fluorescent prot	4.81e-01
16	133	2.1	768	1	W00811	Ras p21 interacting pr	5.64e-01
17	122	2.0	528	1	W62086	Mycobacterium bovis la	3.17e+00
18	122	2.0	528	1	W05831	M. tuberculosis RNA po	3.17e+00
19	122	2.0	530	1	R76480	Virulence-associated s	3.17e+00
20	122	2.0	530	1	W62085	Mycobacterium bovis vl	3.17e+00
21	116	1.9	451	1	W77095	Sorting nexin 2.	7.95e+00
22	117	1.9	485	1	Y00895	Rat GMEB-2' protein se	6.83e+00
23	117	1.9	529	1	Y00894	Rat GMEB-2' protein seq	6.83e+00

24	112	1.8	192	1	W55252	H. pylori ORF 05ep1171	1.45e+01
25	103	1.7	342	1	W55710	H. pylori ORF 06ep1061	5.49e+01
26	104	1.7	371	1	W01113	RNA polymerase primary	4.75e+01
27	104	1.7	382	1	W23969	Connexin protein Cx43.	4.75e+01
28	98	1.6	178	1	R26181	TSG-14.	1.13e+02
29	97	1.6	289	1	W43421	Rat syntaxin 3 protein	1.30e+02
30	102	1.6	449	1	W79140	Plant acetolactate syn	6.35e+01
31	99	1.6	480	1	R77249	Babesia microzote p58.	9.77e+01
32	99	1.6	736	1	W97809	Human GTPase regulator	1.13e+02
33	97	1.6	858	1	R30477	Human leukocyteHGF.	1.30e+02
34	102	1.6	1167	1	W31504	Nematode toxin 167P pr	6.35e+01
35	102	1.6	1167	1	W10653	Bacillus thuringiensis	6.35e+01
36	102	1.6	1168	1	W16326	Nematocidal toxin 167P	6.35e+01
37	98	1.6	2319	1	W53485	Murine factor VIII.	1.13e+02
38	98	1.6	2319	1	W44135	Mus musculus factor VI	1.13e+02
39	100	1.6	2555	1	W81512	Drosophila fat facets	8.47e+01
40	96	1.5	590	1	W94458	Human neurofibromatosis	1.49e+02
41	96	1.5	596	1	W94459	Human neurofibromatosis	1.49e+02
42	96	1.5	740	1	R68838	Plasmodium falciparum	1.49e+02
43	96	1.5	798	1	R57283	Bovine enterokinase.	1.49e+02
44	96	1.5	1086	1	W81839	Human LFA-1 alpha chain	1.49e+02
45	96	1.5	3457	1	W84560	Polypeptide encoded by	1.49e+02

ALIGNMENTS

RESULT	1	
ID	R91597	standard; Protein; 1077 AA.
AC	R91597;	
DT	26-JUL-1996	(first entry)
DE	C3G protein.	
KW	C3G; ras protein guanine nucleotide exchange factor; diagnosis;	
KW	treatment; malignant tumour; activation.	
OS	Homo sapiens.	
PN	J07051073-A.	
PD	28-FEB-1995.	
PF	13-JUN-1994; 130699.	
PR	11-JUN-1993; JP-140806.	
PA	(SHK1) SHINGIJUMSU JICODAN.	
DR	WPI: 95-127357/17.	
DR	N-PSDB; T13347.	
PT	Ras protein guanine nucleotide exchange factor C3G gene - useful for	
PT	diagnosis and treatment of malignant tumours associated with ras	
PT	oncogene activation	
PS	Claim 9; Page 7-8; 9pp; Japanese.	
CC	The present sequence is that of C3G protein which is a ras protein	
CC	guanine nucleotide exchange factor. The gene (T13347) and protein are	
CC	useful for diagnosis or treatment of malignant tumours associated with	
CC	activation of the ras gene.	
SQ	Sequence 1077 AA;	
DB	834 GTLHFSHEIAEQTLTLDALFYKI-DIPEVL-LMKKEQNEKSPNLTQTEHFNNMSY 891	
OY	604 GTFELMSSKDLAYQDTITDWEFLFCVHLELITHTFGHNRKKTATANDLFLRFNFIIQF 663	
DB	892 WRSITIMLOEKAQDERLLFKIKMKRLNNFNSVLAISALDSAPIRLE--WOKQ 949	
OY	664 WVTYICICLSQSKRVQLKKFKIKIAACREKYNLSFALVMGSLNVAIRLITWEKL 723	
DB	950 TSE--GL-AEYCTLIDSSSFRAVRAALSEVEPCIPYLIITLIDTLFVHLGNDPYIDGK 1006	
OY	724 PSKFKFYAEFESLMDPSRNRHRAVRLTVAKLEPLIPMLIKDMFTFHGNTFIDNL 783	
DB	1007 VNEFKRMQFENLDSMRFCQAHYD 1031	
OY	784 VNEFKRMIAVTAIVTAVRYRSQPFN 808	
RESULT	2	

ID R50991 standard; Protein; 489 AA.  
AC R50991;  
DT 11-MAY-1994 (first entry)  
DE Peptide which modulates GDP exchange in the p21-GDP complex.  
KW GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;  
KW detection; cancer; ras; oncogene.  
OS Homo sapiens.  
PN M09321314-A.  
PD 28-OCT-1993.  
PF 19-APR-1993; F00382.  
PR 21-APR-1992; FR-004827.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Schweighoffer F, Tocque B;  
DR WPJ: 93-351724/44.  
DR N-PSDB: 051233.  
PT New peptide(s) modulating GDP exchange in complexes - with ras  
PT protein and derived antibodies, nucleic acid etc. esp. for  
PT diagnosis and treatment of cancer  
PS Claim 2; Page 25-26; 46pp; French.  
CC The peptide antagonises the interaction of GDP-exchange factor (GRF)  
CC with the p21-GDP complex and thereby regulates the activity of ras  
CC gene products. It can also be used to identify other compounds  
CC which can modulate GDP exchange. The coding sequence can be used  
CC to produce antisense products which can inhibit oncogene expression  
CC and for use in diagnosis e.g. detecting (over)expression of GRF for  
CC typing of cancers.  
SQ Sequence 489 AA;

Query Match	4.68;	Score 287;	DB 1;	Length 489;
Best Local Similarity	25.68;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	49;	Mismatches 66; Indels 1; Gaps 1

D0 KHEIDISLILSEIIRINDINIAVARSALEEKWAADICRCHTANYAAVEITISSMRSKSIIFR 359  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY RRFNEIOFWVTEICLSQSLSKRVQLLKFKIKIANHCREKNLNNSFPATVMGSLNIVSR 715  
  
D6 LKKTKMKYSKOTKLIDLKQLYSSEGREFKULREALNCOPCPXYTGMLTLDTAFIEEG 419  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
OY LATWKEIPSKFFKYFAEFESLMPSRRNRHRYRLTWAKLEPLFLFMPLLDKMTFTHEG 775  
  
Db 420 TPNTDEGLVNSCKRMISHIIREIPOQTAYKIE 455  
  
Oy 776 NKTEDN-LVNFEKKMRIMANTARVARYRSOPFMD 810

RESULT	3
ID	R50990 standard; Protein; 666 AA.

DT 11-MAY-1994 (first entry)  
DE Peptide which modulates GDP exchange in the p21-GDP complex.  
KW GDP; exchange; GRF; GDP exchange factor; antitense; diagnosis;  
KW detection; cancer; ras; oncogene.  
OS Homo sapiens.  
PN MO9321314-A.  
PD 28-OCT-1993.  
PF 19-APR-1993; F00382.  
PR 21-APR-1992; FR-004827.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Schweighoffer F, Tocque B;  
DR WPI: 93-351724/44.  
DR N-PSDB; Q51233.  
PT New peptide(s) modulating GDP exchange in complexes - with ras  
PT protein and derived antibodies, nucleic acid etc. esp. for  
PT diagnosis and treatment of cancer  
PS Claim 2; Page 23-24; 46pp; French.  
CC The peptide antagonises the interaction of GDP-exchange factor (GRF)  
CC with the p21-GDP complex and thereby regulates the activity of ras  
CC gene products. It can also be used to identify other compounds  
CC which can modulate GDP exchange. The coding sequence can be used  
CC to produce antisense products which can inhibit oncogene expression  
CC and for use in diagnosis e.g. detecting (over)expression of GRF for  
CC typing of cancers.  
Q0 Sequence 666 AA;

Query Match	4.68;	Score 287;	DB 1;	Length 666;
Best Local Similarity	25.68;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	49;	Mismatches 66;
			Indels	1;
			Gaps	1

[illegible]

RESULT	4
ID	R43578 standard; Protein: 814 AA.
AC	R43578;
DT	11-MAY-1994 (first entry)
DE	Peptide which modulates GDP exchange in the p21-GDP complex.
KW	GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis;
KW	detection; cancer; ras; oncogene.
OS	Homo sapiens.
PN	W09321314-A.
PD	28-OCT-1993.
PF	19-APR-1993; F00382.
PR	21-APR-1992; FR-004827.
PA	(RHON ) RHONE POULENC RORER SA.
PI	Schweighoffer F, Tocque B;
DR	WPI; 93-351724/44.

PT New peptide(s) modulating GDP exchange in complexes - with ras  
PT protein and derived antibodies, nucleic acid etc. esp. for  
PT diagnosis and treatment of cancer  
PS Claim 2; Page 20-22; 46pp; French.  
CC The peptide antagonizes the interaction of GDP-exchange factor (GRF)  
CC with the p21-GDP complex and thereby regulates the activity of ras  
CC gene products. It can also be used to identify other compounds  
CC which can modulate GDP exchange. The coding sequence can be used  
CC to produce antisense products which can inhibit oncogene expression  
CC and for use in diagnosis e.g. detecting (over)expression of GRF for  
CC typing of cancers.  
CC Sequence 814 AA.

Query Match	4.68;	Score 287;	DB 1;	Length 814;
Best Local Similarity	25.68;	Pred. No. 9.88e-13;		
Matches	40;	Conservative	49;	Mismatches 66;
			Indels	1;
			Gaps	1;

```

Db 625 KHNNDISNLJAEIIRNEDINARVSALEKWNVAADICRCLHNTNVALETITSSMRSAIFR 608
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 RRNEIQFWWVTEIELCSQSLRRVQLKKFKILAHCKEYKNLSFFALVWGLSIAMVASR 715
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 685 LKRTWLKVSQKTRALIDKQKYVSEBGRFNKLREALKNDPCVPRLGNYLTDLAFIEEG 744
      | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 716 LALTWELSKRRKFFAEFESLMPDSRNHRAYRRLTVAKLEPLRIPMLLKIDMTFHEG 775
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 745 TPVYEDGLVNSFKRMKMSIIRLROQDQAKIE 780
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 776 NKTJIDN-LVNECKRMKMAITRVRYSQOPFNP 810
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	5
ID	R43580 standard; Protein; 652 AA.
AC	R43580:
DT	11-MAY-1994 (first entry)
DE	Peptide which modulates GDP exchange in the p21-GDP complex.
KW	GDP; exchange; GRF; GDP exchange factor; antisense; diagnosis
KW	detection; cancer; ras; oncogene.
OS	Homo sapiens.
PN	W09321314-A.

PD 28-OCT-1993.  
PF 19-APR-1993; F00382.  
PR 21-APR-1992; PR-004827.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Schweighoffer F, Tocque B;  
DR WPI: 93-351724/44.  
N-PSDB: 051335.  
PT New peptides: modulating GDP exchange in complexes - with ras  
protein and derived antibodies, nucleic acid etc. esp. for  
diagnosis and treatment of cancer  
PS Claim 2: Page 32-34; 46pp; French  
CC The peptide antagonises the interaction of GDP-exchange factor (GRF  
CC with the p21-GDP complex and thereby regulates the activity of ras  
CC gene products. It can also be used to identify other compounds  
CC which can modulate GDP exchange. The coding sequence can be used  
CC to produce antisense products which can inhibit cancer expression  
CC and for use in diagnosis e.g. detecting (over)expression of GRF for  
CC typing of cancers.  
CC Sequence 652 AA;  
SQ

Query Match	4.48;	Score 276;	DB 1;	Length 652;
Best Local Similarity	25.88;	Pred. No. 7.60e-12;		
Matches	59;	Conservative	71;	Mismatches 89;
				Indels 10;
				Gaps 7

[illegible]

RESULT 6  
ID RA7048 standard; Protein; 1297 AA.  
AC RA7048;  
DT 26-MAY-1994 (first entry)  
DE Mammalian son of sevenless (mSos 2) gene product.  
KW Son of sevenless gene; mSos; Sos; guanine exchange factor; Ras;  
KM tyrosine kinase; tumour.  
OS Mus musculus.  
PN M09316179-A.  
PD 19-AUG-1993.  
PE 17-FEB-1993; AU0068.  
PR 17-FEB-1992; AU-000921.  
PA (FLORE-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.  
PI Bowtell DDL;  
DR WPI: 93-272883/34.  
DR N-PSDB: 058624.  
PT DNA encoding mammalian son of seven-less gene and proteins  
PT encoded by it - regulates tyrosine kinase signalling pathway via  
PT ras proteins, may be used for diagnosis and therapy of genetic  
PT defects  
PR Claim 6; Figure 2; 78pp; English.  
CC The mammalian son of sevenless gene has been found to have homology  
CC with various guanine exchange factors, suggesting a possible role in  
CC protein tyrosine kinase pathways via the Ras proteins. Mutations in  
CC the genes and consequently their products may result in specific  
CC defects or tumour formation. Antagonists to the protein/gene may be  
CC used to treat conditions arising from mutations.  
SO Sequence 1297 AA;

query Match	4.38;	Score 266;	DB 1;	Length 1297;
Best Local Similarity	24.98;	Pred. No. 4.81e-11;		
Matches	57;	Conservative	70;	Mismatches 92;
			Indels 10;	Gaps 7

[illegible]

RESULT	7
ID	R84639 standard; Protein; 1297 AA.

DT	25-FEB-1996	(first entry)
DE	msos2 protein.	
KW	msos2; SOS2; Grb2; BCR-ABL; tyrosine kinase; transformation; Ras;	
KW	oncoprotein; leukaemia; Ras guanine nucleotide releasing protein.	
OS	Mus musculus.	
PH	Key	Location/Qualifiers
FT	region	1107..1119
FT		/label= Proline-rich_motif
FT	region	1133..1145
FT		/label= Proline-rich_motif
FT	region	1164..1176
FT		/label= Proline-rich_motif
FT	region	1178..1189
FT		/label= Proline-rich_motif
PM	CA2113494-A.	
PD	15-JUL-1995.	
PF	14-JAN-1994; 113494.	
PR	14-JAN-1994; CA-113494.	
PA	(MOUN ) MOUNT SINAI HOSPITAL CORP.	
PA	(TEXA ) UNIV TEXAS.	
PI	Arlinghaus R, Gish G, Liu J, Pawson A, Pull L;	
PI	WPI; 95-302931/40.	
DR	N-PSDB; T05114.	
PT	Detection of agents that modify BCR-ABL mediated transformation -	
PT	useful in treatment of leukaemia and other malignancies	
PS	Dislosure; Page 62-67; 10pp; English.	
CC	Mouse msos2 (R84639) is a Ras quantine nucleotide releasing protein.	
CC	The SH3 domains of Grb2 associate the the proline-rich domains of	
CC	msos2 and msos1 (R84638). Synthetic peptides (R84640-48) based on	
CC	these motifs can be used to screen for compounds that affect BCL-ABL	
CC	mediated transformation. Such compounds have value in the treatment	
CC	of chronic, acute myelogenous or acute lymphocytic leukaemia.	
SQ	Sequence 1297 AA;	

Query Match	4.28;	Score 262;	DB 1;	Length 1297;
Best Local Similarity	24.98;	Pred. No. 1.00e-10;		
Matches 57; Conservative		70; Mismatches 92;	Indels 10;	Gaps 7;

```

Db 717 FESSPPVEHNTSRGQFETDPMTHIETARLOITLSDYRKOPELSGVWTKRED 776
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 566 FDSLTPRPEDEGFTYGVSTPEFLMS--SKDLYMTITYMELFCYNHELELYHTEGRHN 643
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 777 KEINSPNLMKIMHTNTLTLMFECIKYBOKNFEEBAVUSRIYELQVFODLNNNGYLE 836
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 644 FKTTATLDTLFRRENEIDFMWVTEICLCSOLSKRYOLLKFKFLIAHCKEKNINSEFA 703
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 837 IVATVNSVSVSYRDLHFEFLAOLR--KRRLD--DA--VELSDPHEFKYLVKLSINPCEVFF 893
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 704 IYMLSLINATSRALTWELBPKFKFFAHEESLMDSNRHN--RAKRLYIAKLEPPLIDEM 766
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 894 GIYTLNLTKEGNSDFLKRKGKDLINFSKRKYVAETGTGEOYQNOPRY 942
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```





CC otherwise ameliorate, the effects of a cancer in an animal or bird.  
Sequence 728 AA;

Query Match 2.9%; Score 180; DB 1; Length 728;  
Best Local Similarity 26.4%; Pred. No. 2,296-04;

Matches 38; Conservative 34; Mismatches 68; Indels 4; Gaps 4;

DB 269 FDLHPEMELAEHLTYLEYSR-C-KILFODYHSPVTCVDPVLERFISFNYSOVN 326

QY 606 FELMSKDLAYOMTYIDWELFNCVHELELYHTGRNNEKKTANDELRLRPNELQFVY 665

DB 327 QLMY-LSKPTAPORALVTHFVHAELKLOLONFTLMAVGGSLSSSRLEKTHSVS 385

QY 666 VTEICLCOLS-KRVLCKRKFIKAHCKEYKLNLSFFAIVMGLSNIAVRLATWEKLP 724

DB 386 PETIKIMEGLTELVATGNYCNRY 409

QY 725 SKKFFYAEFESIMDPSRNHRYR 748

# RESULT 14

ID W85040 standard; Protein; 932 AA.

AC W85040;

DT 08-FEB-1999 (first entry)

DE PKG-green fluorescent protein fusion product.

KW Human; PKG gene; fusion protein; green fluorescent protein; GFP;

OS Intracellular signalling; chimera.

OS Chimeric - Aequorea victoria.

PN W09845704-A2.

PD 15-OCT-1998.

PF 07-APR-1998; DK0145.

PR 07-APR-1997; DK-000392.

PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,

PI Tullin S;

DR WPI; 98-594491/50.

DR N-PSDB; V71085.

PT Determining effect on signalling pathways in live cells from

PT redistribution of luminoephores - specifically fusions of green

PT fluorescent protein with a signalling component and new apparatus,

PT particularly for identifying toxins and potential therapeutic agents

PS Example 10; Pages 281-283; 326pp; English.

CC The present sequence represents a human PKG-green fluorescent protein

CC (GFP) fusion protein. The fusion protein is used in an assay to

CC exemplify the invention. The specification describes how quantitative

CC information about the influence of a molecule on a cellular response is

CC obtained by recording the variation, caused by the molecule, on

CC mechanically intact living cells, in the spatially distributed light

CC emitted from a luminoephore present in the cells. The variation in light

CC distribution to the degree of the molecule. The method is used to

CC identify agents that (in)directly affect intracellular signalling,

CC especially to screen for potential therapeutic agents or toxins, and to

CC identify new drug targets.

Sequence 932 AA;

Query Match 2.2%; Score 134; DB 1; Length 932;  
Best Local Similarity 30.6%; Pred. No. 4,816-01;

Matches 37; Conservative 28; Mismatches 46; Indels 10; Gaps 8;

DB 227 YMEFLKVPFOSLPBEILSKLADYLE-ETHYENGXYIIROGARGDPFFIISGTVNTR 285

QY 186 YEBLLH-ITALSHLSTYKRELAVGLIFESHAKGTVLFGNGEGTSWYIILGVSNNVI 244

DB 286 EDSPSEDVFLRTLGKDMFGEKALOGEDVRTANVIAEAATC--LVYIDRSFKHLIGL 343

QY 245 --Y-GKG-V-VCTLHGGDFGKALVNDAPRAASIVLRD-NCHFLVNDKEDFNRIIRDV 298

DB 344 D 344

QY 299 E 299

RESULT 15  
ID W85039 standard; Protein; 933 AA.

AC W85039;

DT 08-FEB-1999 (first entry)

DE Green fluorescent protein-PKG fusion product.

KW Human; PKG gene; fusion protein; green fluorescent protein; GFP;

OS Intracellular signalling; chimera.

OS Chimeric - Aequorea victoria.

PN W09845704-A2.

PD 15-OCT-1998.

PF 07-APR-1998; DK0145.

PR 07-APR-1997; DK-000392.

PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,

PI Tullin S;

DR WPI; 98-594491/50.

DR N-PSDB; V71084.

PT Determining effect on signalling pathways in live cells from

PT redistribution of luminoephores - specifically fusions of green

PT fluorescent protein with a signalling component, and new apparatus,

PT particularly for identifying toxins and potential therapeutic agents

PS Example 10; Pages 274-276; 326pp; English.

CC The present sequence represents a green fluorescent protein (GFP)-human

CC PKG fusion protein. The fusion protein is used in an assay to

CC exemplify the invention. The specification describes how quantitative

CC information about the influence of a molecule on a cellular response is

CC obtained by recording the variation, caused by the molecule, on

CC mechanically intact living cells, in the spatially distributed light

CC emitted from a luminoephore present in the cells. The variation in light

CC emission is processed to provide information that correlates spatial

CC distribution to the degree of the molecule. The method is used to

CC identify agents that (in)directly affect intracellular signalling,

CC especially to screen for potential therapeutic agents or toxins, and to

CC identify new drug targets.

Sequence 933 AA;

Query Match 2.2%; Score 134; DB 1; Length 933;  
Best Local Similarity 30.6%; Pred. No. 4,816-01;

Matches 37; Conservative 28; Mismatches 46; Indels 10; Gaps 8;

DB 474 YMEFLKVPFOSLPBEILSKLADYLE-ETHYENGXYIIROGARGDPFFIISGTVNTR 532

QY 186 YEBLLH-ITALSHLSTYKRELAVGLIFESHAKGTVLFGNGEGTSWYIILGVSNNVI 244

DB 533 EDSPSEDVFLRTLGKDMFGEKALOGEDVRTANVIAEAATC--LVYIDRSFKHLIGL 590

QY 245 --Y-GKG-V-VCTLHGGDFGKALVNDAPRAASIVLRD-NCHFLVNDKEDFNRIIRDV 298

DB 591 D 591

QY 299 E 299

Search completed: Sat Apr 22 13:18:22 2000  
Job time : 49 secs.